

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

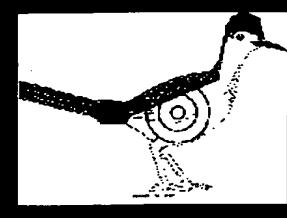


FIG. 1

1/40

File Edit View Go Communicator Help

Bookmarks Location http://victoria.inpharmatica.co.uk/~volker/BPD3target.html

 Target Mining Interface



**Select Your Query Sequence**

• Enter PDB accession number (e.g. 1QMA):  1ifa and chain (e.g. B):  a

OR

• Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

**Select Database**

Release:  DEVF9 = BPD3

**Apply Filters**

• Iteration Filter: PSI-BLAST matches to be excluded:  
 Matches detected during the first 20 forward iterations

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first\_PB\_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

• Filter for the following SPECIES:

Homo sapiens  Rattus norvegicus (Rat)  Mus musculus (Mouse)  Danio rerio (Zebra fish)

100%

FIG. 2A

2/40

File Edit View Go Communicator  
Bookmarks Location http://London-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl Help

2) 84 additional hits identified by both, Genome Threader and PSI-BLAST:

Combined Genome Threader and PSI-Blast output: PSI-BLAST values are shown in maroon!

Add2list	BPD links	WWW link	Title	Organism	Div.	%ID (GT,PSI)	Query rgn. (GT,PSI)	Target rgn. (GT,PSI)	Ah. score (GT)	Conf. (GT)	1st Iter. (PSI)	Best Iter. (PSI)	E-value (PSI)
↓	AAA59544.1 drill through Top50BlastHits	<a href="#">AAA59544.1</a>	Not given	Homo sapiens	PRI	33.7%, 33% unmaskedSW	4-183, 4-183	150-336, 150-336	449	100% unmaskedGT	1	2	3E-
→	AAB24821.1 drill through Top50BlastHits	<a href="#">AAB24821.1</a>	leukocyte integrin alpha chain	Homo sapiens	PRI	33.7%, 33% unmaskedSW	4-183, 4-183	150-336, 150-336	449	100% unmaskedGT	1	2	3E-
↓	Q99715 drill through Top50BlastHits	<a href="#">Q99715</a>	COLLAGEN ALPHA I(XII) CHAIN PRECURSOR.	Homo sapiens (Human)	PRI	28.8%, 26% unmaskedSW	4-181, 2-174	140-318, 2321-2495	440	100% unmaskedGT	1	3	2E-
↓	P20701 drill through Top50BlastHits	<a href="#">P20701</a>	LEUKOCYTE ADHESION GLYCOPROTEIN LFA-1 ALPHA CHAIN PRECURSOR (LEUKOCYTE FUNCTION ASSOCIATED MOLECULE 1, ALPHA CHAIN)(CD11a) (INTEGRIN ALPHA-L).	Homo sapiens (Human)	PRI	100%, 100% unmaskedSW	1-183, 1-183	153-335, 153-335	423	100% unmaskedGT	1	1	1E-
↓	AAC31672.1 drill through Top50BlastHits	<a href="#">AAC31672.1</a>	leukocyte function-associated molecule-1 alpha subunit	Homo sapiens	PRI	99.5%, 99% unmaskedSW	1-183, 1-183	153-335, 153-335	423	100% unmaskedGT	1	1	1E-
↓	CAA72402.1 drill through Top50BlastHits	<a href="#">CAA72402.1</a>	collagen type XIV	Homo sapiens	PRI	29.1%, 29% unmaskedSW	2-180, 2-180	5-185, 5-185	422	100% unmaskedGT	1	2	2E-
↓	AAB38702.1 drill through Top50BlastHits	<a href="#">AAB38702.1</a>	cartilage matrix protein	Homo sapiens	PRI	31.7%, 27% unmaskedSW	4-183, 2-182	275-455, 39-223	413	100% unmaskedGT	1	2	2E-
↓	CAB70853.1 drill through Top50BlastHits	<a href="#">CAB70853.1</a>	hypothetical protein	Homo sapiens	PRI	28%, 28% unmaskedSW	1-180, 1-183	437-620, 437-624	406	100% unmaskedGT	1	2	1E-
↓	CAA27972.1 drill through Top50BlastHits	<a href="#">CAA27972.1</a>	Not given	Homo sapiens	PRI	20.5%, 20% unmaskedSW	3-183, 2-181	1497-1673, 1689-1873	405	100% unmaskedGT	2	3	1E-
↓	AAB59512.1 drill through Top50BlastHits	<a href="#">AAB59512.1</a>	Not given	Homo sapiens	PRI	20.5%, 20% unmaskedSW	3-183, 2-181	758-934, 950-1134	405	100% unmaskedGT	2	3	1E-
↓	CAA07569.1 drill through Top50BlastHits	<a href="#">CAA07569.1</a>	matriñ-4	Homo sapiens	PRI	28.1%, 25% unmaskedSW	1-183, 1-183	342-528, 31-217	403	100% unmaskedGT	1	2	2E-

Entrez	Redundant sequence display	BPD link	WWW link	Title	Organism	Div.	% ID	Query frag.	Target frag.	Aln. score	Conf.
J	<u>Red Seq. View</u> drill through Top50BlastHits	<u>AAE71133.1</u>	<u>AAE71133.1</u>	<u>PRQ2769</u>	<u>Homo sapiens</u>	<u>PRI</u>	<u>13.8% unmaskedSW</u>	<u>109-179</u>	<u>1-80</u>	<u>122</u>	<u>100% unmaskedGT</u>
J	<u>Red Seq. View</u> drill through Top50BlastHits	<u>CAE52192.1</u>	<u>CAE52192.1</u>	<u>G7c protein</u>	<u>Homo sapiens</u>	<u>PRI</u>	<u>9.6% unmaskedSW</u>	<u>7-117</u>	<u>20-124</u>	<u>82</u>	<u>99.18% unmaskedGT</u>
J	<u>Red Seq. View</u> drill through Top50BlastHits	<u>CAA82910.1</u>	<u>CAA82910.1</u>	<u>basic transcription factor 2, 44 kD subunit</u>	<u>Homo sapiens</u>	<u>PRI</u>	<u>11.2% unmaskedSW</u>	<u>5-164</u>	<u>61-225</u>	<u>75</u>	<u>99.18% unmaskedGT</u>
J	<u>Red Seq. View</u> drill through Top50BlastHits	<u>AAQ21820.1</u>	<u>AAQ21820.1</u>	<u>NC37</u>	<u>Homo sapiens</u>	<u>PRI</u>	<u>9.6% unmaskedSW</u>	<u>7-117</u>	<u>318-422</u>	<u>82</u>	<u>98.86% unmaskedGT</u>
J	<u>Red Seq. View</u> drill through Top50BlastHits	<u>BA20761.1</u>	<u>BA20761.1</u>	<u>Not given</u>	<u>Homo sapiens</u>	<u>PRI</u>	<u>16.4% unmaskedSW</u>	<u>5-114</u>	<u>1836-1950</u>	<u>78</u>	<u>98.85% unmaskedGT</u>
J	<u>Red Seq. View</u> drill through Top50BlastHits	<u>AAA36154.1</u>	<u>AAA36154.1</u>	<u>Not given</u>	<u>Homo sapiens</u>	<u>PRI</u>	<u>13.2% unmaskedSW</u>	<u>5-112</u>	<u>10-137</u>	<u>79</u>	<u>98.18% unmaskedGT</u>
J	<u>Red Seq. View</u> drill through Top50BlastHits	<u>AAFD03046.1</u>	<u>AAFD03046.1</u>	<u>candidate tumor suppressor protein DCE1</u>	<u>Homo sapiens</u>	<u>PRI</u>	<u>13.8% unmaskedSW</u>	<u>5-113</u>	<u>4-131</u>	<u>79</u>	<u>97.07% unmaskedGT</u>
J	<u>Red Seq. View</u> drill through Top50BlastHits	<u>AAE74854.1</u>	<u>AAE74854.1</u>	<u>orf, hypothetical protein</u>	<u>Escherichia coli</u>	<u>BCT</u>	<u>16.7% unmaskedSW</u>	<u>5-99</u>	<u>250-335</u>	<u>78</u>	<u>95.09% unmaskedGT</u>
J	<u>Red Seq. View</u> drill through Top50BlastHits	<u>AAB60942.1</u>	<u>AAB60942.1</u>	<u>breast cancer suppressor candidate 1</u>	<u>Homo sapiens</u>	<u>PRI</u>	<u>16.2% unmaskedSW</u>	<u>63-156</u>	<u>2-90</u>	<u>80</u>	<u>93.33% unmaskedGT</u>
J	<u>Red Seq. View</u> drill through Top50BlastHits	<u>AAAG67537.1</u>	<u>AAAG67537.1</u>	<u>glycoprotein IIIa</u>	<u>Homo sapiens</u>	<u>PRI</u>	<u>18.9% unmaskedSW</u>	<u>3-112</u>	<u>111-251</u>	<u>72</u>	<u>92.08% unmaskedGT</u>
J	<u>Red Seq. View</u> drill through Top50BlastHits	<u>AAAS5589.1</u>	<u>AAAS5589.1</u>	<u>Not given</u>	<u>Homo sapiens</u>	<u>PRI</u>	<u>19.6% unmaskedSW</u>	<u>3-112</u>	<u>138-277</u>	<u>74</u>	<u>91.34% unmaskedGT</u>
J	<u>Red Seq. View</u> drill through Top50BlastHits	<u>AAAB71380.1</u>	<u>AAAB71380.1</u>	<u>platelet membrane glycoprotein IIIa beta subunit</u>	<u>Homo sapiens</u>	<u>PRI</u>	<u>19.6% unmaskedSW</u>	<u>3-112</u>	<u>138-277</u>	<u>74</u>	<u>91.34% unmaskedGT</u>
J	<u>Red Seq. View</u> drill through Top50BlastHits	<u>AAAAS5927.1</u>	<u>AAAAS5927.1</u>	<u>Not given</u>	<u>Homo sapiens</u>	<u>PRI</u>	<u>19.6% unmaskedSW</u>	<u>3-112</u>	<u>138-277</u>	<u>74</u>	<u>91.34% unmaskedGT</u>

FIG. 2B  
1) 509 hits identified by Genome Threader only:





FIG. 3

FIG. 4

**Pfam**  
Protein families database of alignments and HMMs

[Home](#) | [Keyword search](#) | [Protein search](#) | [DNA search](#) | [Browse Pfam](#) | [Taxonomy search](#) | [Help](#)



**Results for gi|1788084|gb|AAC74854.1|**

There were no matches to Pfam-A (including borderline matches) for gi|1788084|gb|AAC74854.1|

**Matches to Pfam-B**

Domain	Start	End	Evalue	Alignment
Pfam-B_39416	233	423	3.7e-103	<a href="#">Align</a>

[427 residues]

---

**Alignments of Pfam-B domains to best-matching Pfam-B sequence**

Format for fetching alignments to Pfam-B families  Hypertext linked to swisspfpam

Query gi|1788084|gb|AAC74854.1|/233-423 matching [Pfam-B\\_39416](#)

```

YEAM_ECOLI 233 DLRYKNEYKRPDPSSQAVMFCLMDVSGSMDQSTKDMAKRPFYILLYLFLSR 282
          DLRYKNEYKRPDPSSQAVMFCLMDVSGSMDQSTKDMAKRPFYILLYLFLSR
gi|1788084|gb|AAC74854.1| 233 DLRYKNEYKRPDPSSQAVMFCLMDVSGSMDQSTKDMAKRPFYILLYLFLSR 282

YEAM_ECOLI 283 TYKNVEVVYIRMMTQAKEVDEMEFFYSQETGGTIVSSALKLMDDEVVKERY 332
          TYKNVEVVYIRMMTQAKEVDEMEFFYSQETGGTIVSSALKLMDDEVVKERY
gi|1788084|gb|AAC74854.1| 283 TYKNVEVVYIRMMTQAKEVDEMEFFYSQETGGTIVSSALKLMDDEVVKERY 332

YEAM_ECOLI 333 NPAQWNIIYARQASDGDNWADDSPLCMEILAKKLLPVVRYYSYIEITRRAM 382
          NPAQWNIIYARQASDGDNWADDSPLCMEILAKKLLPVVRYYSYIEITRRAM
gi|1788084|gb|AAC74854.1| 333 NPAQWNIIYARQASDGDNWADDSPLCMEILAKKLLPVVRYYSYIEITRRAM 382

YEAM_ECOLI 383 QTLWREYEMLQSTFDNFAMQMIRDQDDIYPVFRELPMKQNA 423
          QTLWREYEMLQSTFDNFAMQMIRDQDDIYPVFRELPMKQNA
gi|1788084|gb|AAC74854.1| 383 QTLWREYEMLQSTFDNFAMQMIRDQDDIYPVFRELPMKQNA 423

```

[Align to family](#)

---

If you think there is anything wrong with this script, please contact [Pfam](#)

FIG. 5

7/40

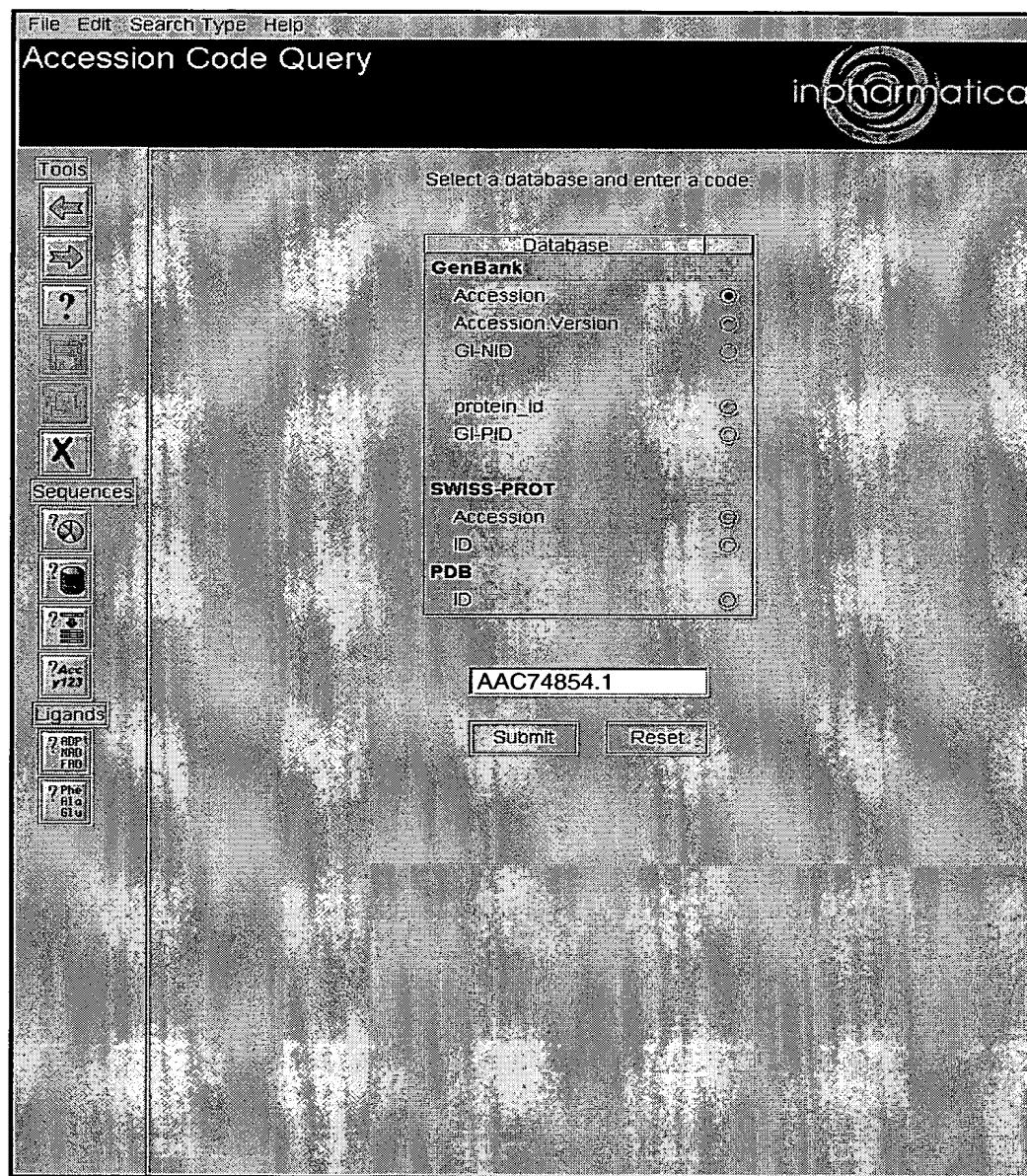
File Edit View Go Communicator Help

Bookmarks Location: <http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=...>

LOCUS AAC74854 427 aa BCT 01-DEC-2000  
 DEFINITION orf, hypothetical protein [Escherichia coli K12].  
 ACCESSION AAC74854  
 PID g1788084  
 VERSION AAC74854.1 GI:1788084  
 DBSOURCE locus AE000273 accession AE000273.1  
 KEYWORDS  
 SOURCE Escherichia coli K12.  
 ORGANISM Escherichia coli K12.  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.  
 1 (residues 1 to 427)  
 Blattner, F.R., Plunkett, G. III, Bloch, C.R., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, M.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.  
 TITLE The complete genome sequence of Escherichia coli K-12  
 JOURNAL Science 277 (5331), 1453-1474 (1997)  
 MEDLINE 97426617  
 PUBMED 9278503  
 2 (residues 1 to 427)  
 Blattner, F.R.  
 REFERENCE AUTHORS Direct Submission  
 TITLE Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
 JOURNAL Email: [ecoli@genetics.wisc.edu](mailto:ecoli@genetics.wisc.edu) Phone: 608-262-2534 Fax: 608-263-7459  
 3 (residues 1 to 427)  
 Blattner, F.R.  
 REFERENCE AUTHORS Direct Submission  
 TITLE Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
 JOURNAL Email: [ecoli@genetics.wisc.edu](mailto:ecoli@genetics.wisc.edu) Phone: 608-262-2534 Fax: 608-263-7459  
 4 (residues 1 to 427)  
 Plunkett, G. III.  
 REFERENCE AUTHORS Direct Submission  
 TITLE Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
 JOURNAL COMMENT This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIM grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA 30332 (e-mail: [mark@amber.gatech.edu](mailto:mark@amber.gatech.edu)). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). \*\*\* The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.  
 FEATURES Method: conceptual translation.  
 source Location/Qualifiers  
 1..427  
 /organism="Escherichia coli K12"  
 /strain="K12"  
 /sub\_strain="MG1655"  
 /db\_xref="taxon:83333"  
 Protein 1..427  
 /function="orf; unknown"  
 /product="orf, hypothetical protein"  
 CDS 1..427  
 /gene="yehA"  
 /coded\_by="1788078:6385..7668"  
 /transl\_table=11  
 /note="0427: This 427 aa ORF is 28 pct identical (43 gaps) to 327 residues of an approx. 312 aa protein YZDC\_BACSU SW: P45742"  
 ORIGIN 1 mtwfidrrln gknksmvnrq rflrrykaqi kqsisearink rsvtdvdsgs svsiptedis  
 61 epmfhggrrgg lrhrvhpgnd hfvgndrier pggggggsgs gggqasqdge ggdefvfqis  
 121 kdeylldlifc dalapnkqm qqrqlteykt hrqytagv panisvvrsl qnslarrtam  
 181 taackrrelha leenlaiisn sepaqlleee rlrkeiaealr akiervpfid tdfdrlyknye

100%

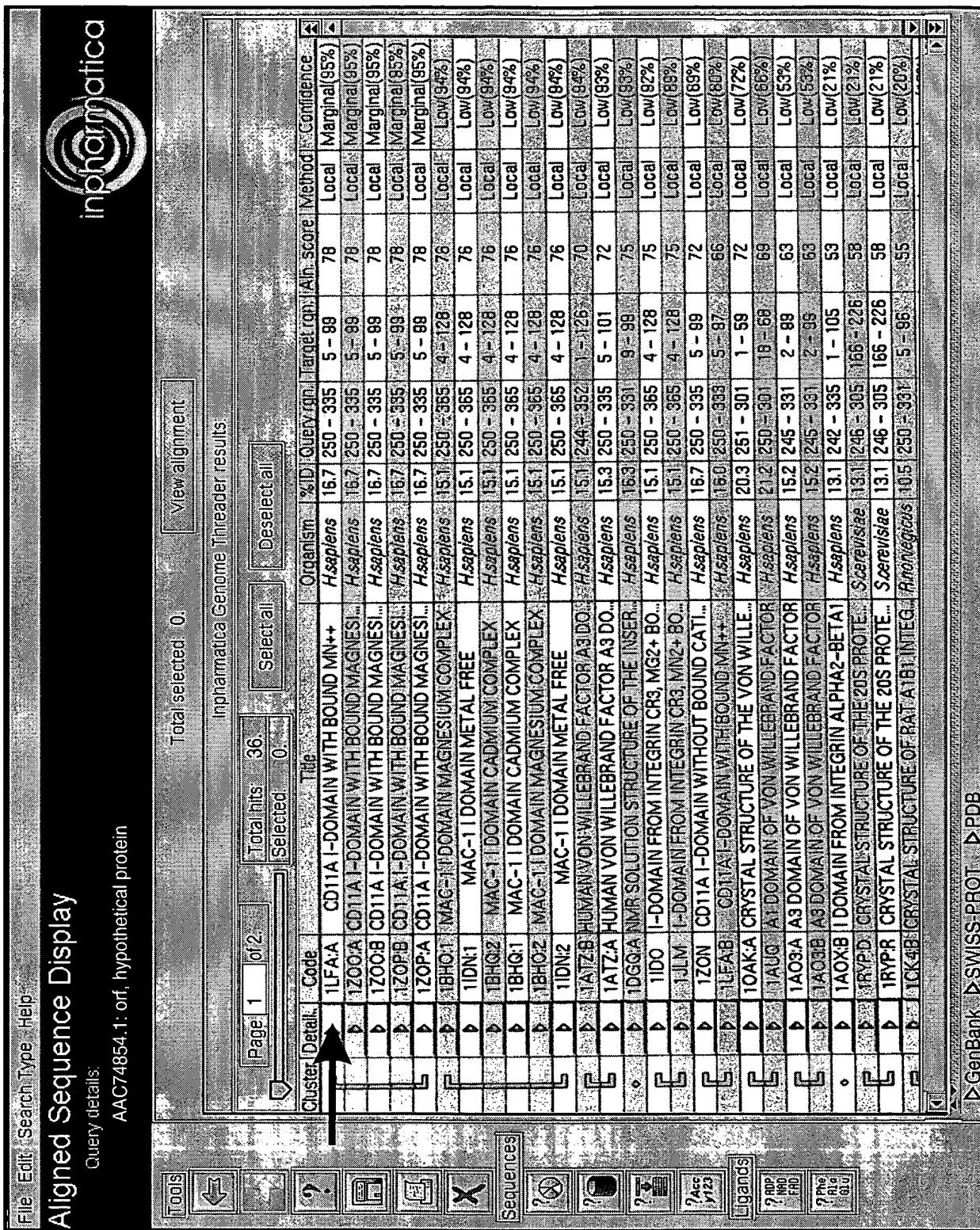
FIG. 6A



**FIG. 6B** Aligned Sequence Display

Query details: AAC74854 1: off hypothetical protein

AAAC74854 1. oft hypothetical protein



inpharmatica

Query details:  
AAC74854.1: orf, hypothetical protein

Total selected: 0

Reverse Maximised PSI-BLAST results:

Cluster Detail	Code	Title	Organism	% ID	Query	Target	Bit score	E value
•	CAB12726.1	Not given	<i>B. subtilis</i>	23.0	14 - 423	21 - 385	1	2.00E-29
•	AAE8986.10.1	Conserved hypothetical protein	<i>Mammouth</i>	15.0	234 - 371	239 - 401	-6	4.00E-08

Sequences

Tools:

► GenBank ► SWISS-PROT ► PDB

FIG. 6C Aligned Sequence Display

FIG. 7  
AIEye output (January 4, 2002 3:07 PM)

11/40

BAA15585.1 MTWFIDRARLNGKNKSMVNQRFLRRYKAQIKQSISEAINKRSVTDVDSGESVSISSPTEDISEPMFHQGRGGLRRHHRVHPGN  
1LFA:A

BAA15585.1 DHFVQNDRIERPQGGGGSGGGQGQASQDGEGQDEFVfqISKDEYLDLFLFEDLALPNLKQNQQRQLTEYKTHRAYTAN  
1LFA:A

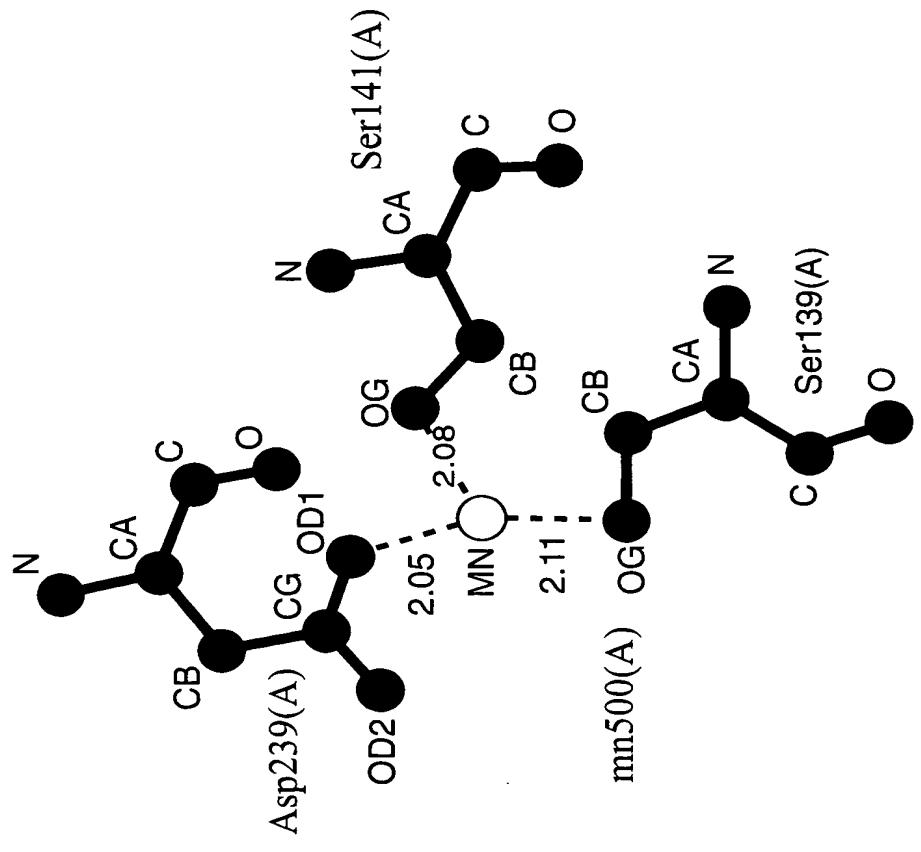
BAA15585.1 GVPANISVVRSLQNSLARRITAMTAGKRAELHALENLAIISNEPAQLEEEERLKEIAELRAKIERVPPFIOTFDLRYK  
1LFA:A

BAA15585.1 NYEKRPDPOSSQAVMFCLMDVSGSMQDQSTKDMAKRFYILLYLFLSRATYKNVEVV  
1LFA:A

BAA15585.1 FYSQETGGTIVSSALKLMDDEVVKERYNPAQWNYYAAQASDGDNWADDSPLCHEILAKKLLPVVRYSSYEITRAA-HQT  
1LFA:A

BAA15585.1 LWREYEHLLQSTFDNFAMQHIRDQDDIYPVFRELHKQNATAKG  
1LFA:A

FIG. 8A



11fa: MN500 Chain [A]

FIG. 8B

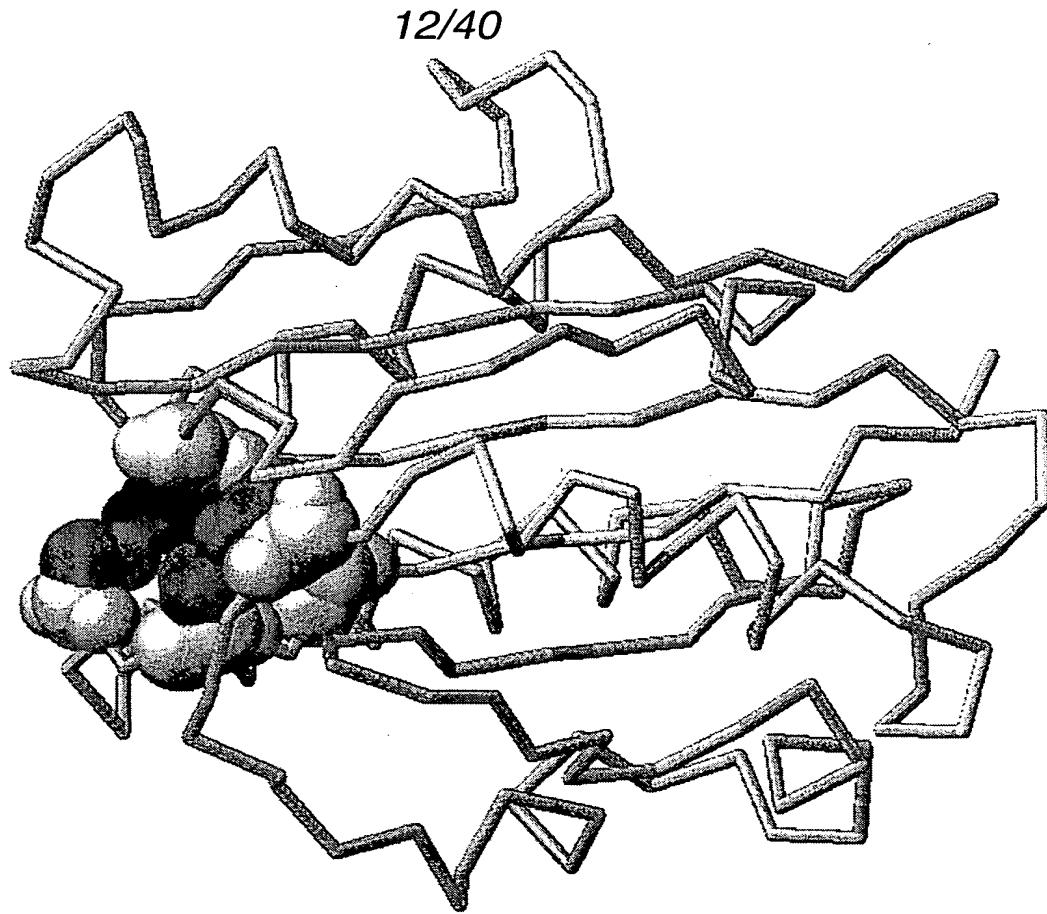
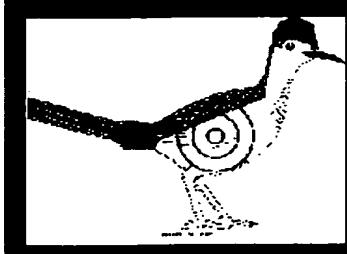


FIG. 9

13/40

File Edit View Go Communicator Help

Bookmarks Location http://victoria.inpharmatica.co.uk/~volker/BPD3target.html

 Target Mining Interface 

**Select Your Query Sequence**

- Enter PDB accession number (e.g. 1QMA):  and chain (e.g. B):

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

**Select Database**

Release:

**Apply Filters**

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first\_PB\_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

FIG. 10A

Combined Genome Threader and PSI - Blast output! PSI - BLAST values are shown in maroon!

2) 82 additional hits identified by both, Genome Threader and PSI - BLAST:

Accession	Ref.Clinic.	WMTW.line:	Title	Organism	Env.	1st Iter. (GT/PSI)	Conf. (GT)	1st Iter. (PSI)	Best Iter. (PSI)
P56199	drillthrough	P26199	INTEGRIN ALPHAI (LAMININ AND COLLAGEN RECEPTOR) (VLA-1) (CD49a).	Homo sapiens (Human).	PRI	51.8% 51% unmasked SW	2-200, 2-200	139-337, 487	100% unmasked GT
-	drillthrough	P17301	PLATELET MEMBRANE GLYCOPROTEIN IA PRECURSOR (GP1A) (COLLAGEN RECEPTOR) (INTEGRIN ALPHAI-2) (VLA-2 ALPHA CHAIN) (CD49b).	Homo sapiens (Human).	PRI	98.5% 10% unmasked SW	2-200, 2-200	169-367, 471	100% unmasked GT
-	drillthrough	AA010291_1	Integrin alpha-11 subunit precursor	Homo sapiens	PRI	45.2% 45% unmasked SW	2-200, 2-200	159-355, 469	100% unmasked GT
-	drillthrough	AA010291_2	Integrin alpha 11 subunit precursor	Homo sapiens	PRI	45.2% 45% unmasked SW	2-200, 2-200	159-355, 469	100% unmasked GT
-	drillthrough	Q9215	COLLAGEN ALPHA (XII) CHAIN PRECURSOR.	Homo sapiens (Human).	PRI	28.1% 31% unmasked SW	2-196, 7-189	133-316, 437	100% unmasked GT
-	drillthrough	AA051919_2	No given	Homo sapiens	PRI	27.1% 27% unmasked SW	2-200, 2-200	144-338, 435	100% unmasked GT
-	drillthrough	AA051919_1	No given	Homo sapiens	PRI	27.1% 27% unmasked SW	2-200, 2-200	144-338, 435	100% unmasked GT
-	drillthrough	AA052341_1	No given	Homo sapiens	PRI	27.1% 27% unmasked SW	2-200, 2-200	144-338, 435	100% unmasked GT
-	drillthrough	AA052341_2	No given	Homo sapiens	PRI	27.1% 27% unmasked SW	2-200, 2-200	144-338, 435	100% unmasked GT
-	drillthrough	CA971222_1	d1238D15.1 (collagen, type XII, alpha 1)	Homo sapiens	PRI	26.9% 26% unmasked SW	5-195, 156-201	233-472, 424	100% unmasked GT
-	drillthrough	P21941	CARTILAGE MATRIX PROTEIN PRECURSOR (MATRILIN-1).	Homo sapiens (Human).	PRI	26.8% 26% unmasked SW	3-198, 3-198	271-555, 37-224	100% unmasked GT
-	drillthrough	AA010506_1	Type XII collagen	Homo sapiens	PRI	22.6% 30% unmasked SW	2-195, 2-195	133-318, 422	100% unmasked GT
-	drillthrough	AA010506_1	Type XII collagen	Homo sapiens	PRI	22.6% 30% unmasked SW	2-195, 2-195	133-318, 422	100% unmasked GT
-	drillthrough	AA052342_1	Integrin subunit alpha 10 precursor	Homo sapiens	PRI	48.5% 48% unmasked SW	2-199, 2-199	162-359, 419	100% unmasked GT
-	drillthrough	CA007591_1	matrillin-4	Homo sapiens	PRI	26.5% 26% unmasked SW	1-199, 2-200	335-529, 411	100% unmasked GT

FIG. 10B

15/40

608 out of these 632 PSI-BLAST matches were identified using 'positive iterations'

FIG. 10C

16/40

24 out of these 632 PSI-BLAST matches were identified using 'negative iterations'

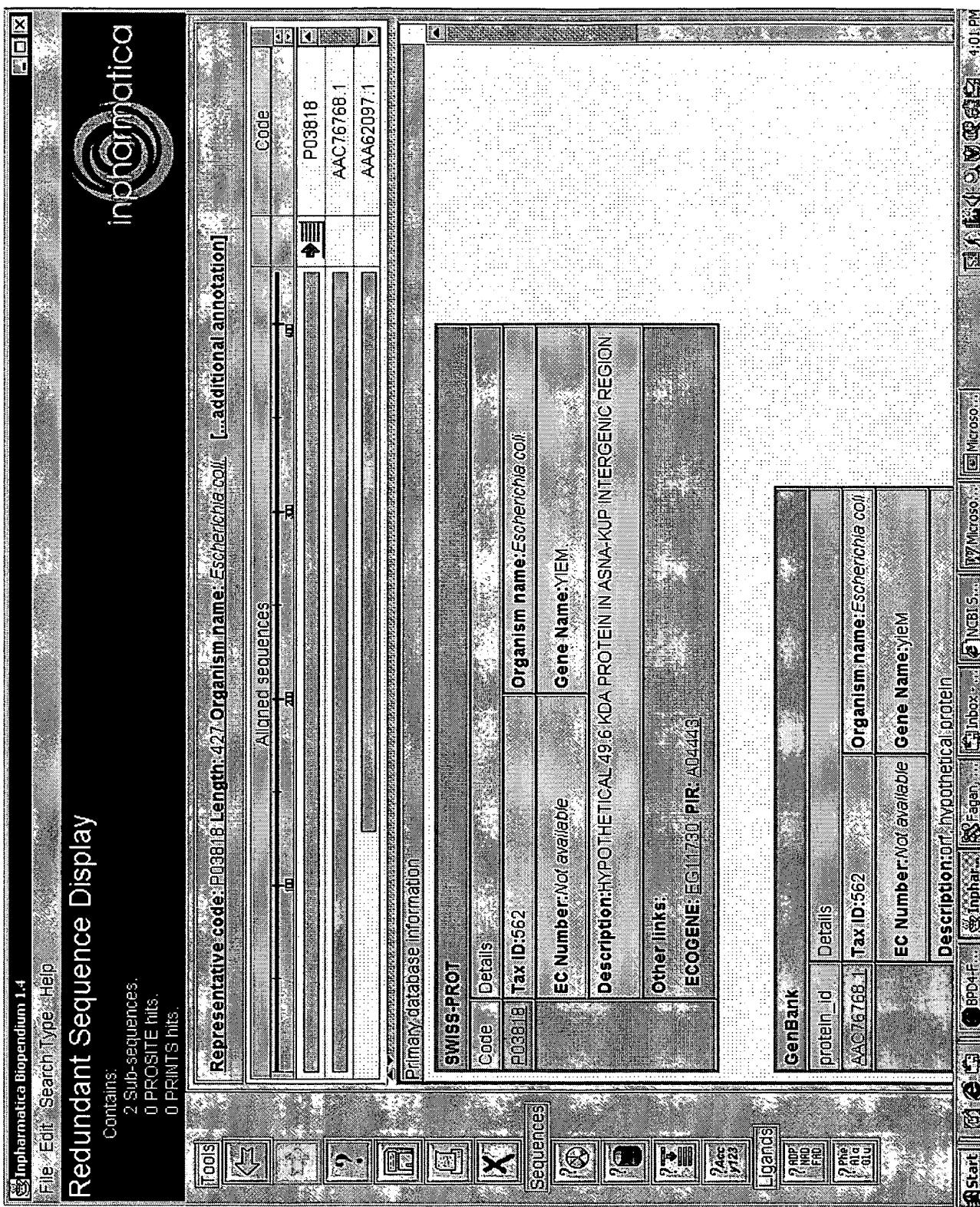


FIG.

FIG. 12

**Pfam**  
Protein families database of alignments and HMMs

[Home](#) | [Keyword search](#) | [Protein search](#) | [DNA search](#) | [Browse Pfam](#) | [Taxonomy search](#) | [Help](#)



**Results for gi|2367274|gb|AAC76768.1|**

There were no matches to Pfam-A (including borderline matches) for gi|2367274|gb|AAC76768.1|

**Matches to Pfam-B**

Domain	Start	End	Evalue	Alignment
Pfam-B_1S204	204	408	2.4e-108	<a href="#">Align</a>

[427 residues]

**Alignments of Pfam-B domains to best-matching Pfam-B sequence**

Format for fetching alignments to Pfam-B families:  Hypertext linked to swisspfa

Query gi|2367274|gb|AAC76768.1|/204-408 matching Pfam-B\_1S204

```

YIEV_ECOLI 204 DILRLPPPELTLGITELEYEFYRRLVEKQLLTYRLNGESWREKVIERPV 253
          DILRLPPPELTLGITELEYEFYRRLVEKQLLTYRLNGESWREKVIERPV
gi|2367274|gb|AAC76768.1| 204 DILRLPPPELTLGITELEYEFYRRLVEKQLLTYRLNGESWREKVIERPV 253

YIEV_ECOLI 254 VMDYDEQPRGPFIVCVDTSGSM66FREQCAKAFCLALMRLALRENRCY 303
          VMDYDEQPRGPFIVCVDTSGSM66FREQCAKAFCLALMRLALRENRCY
gi|2367274|gb|AAC76768.1| 254 VMDYDEQPRGPFIVCVDTSGSM66FREQCAKAFCLALMRLALRENRCY 303

YIEV_ECOLI 304 IMLPSTEIVRYTELSGPQ6IEQAIRFLSQQPRGGTDLASCPRAIMERLQSR 353
          IMLPSTEIVRYTELSGPQ6IEQAIRFLSQQPRGGTDLASCPRAIMERLQSR
gi|2367274|gb|AAC76768.1| 304 IMLPSTEIVRYTELSGPQ6IEQAIRFLSQQPRGGTDLASCPRAIMERLQSR 353

YIEV_ECOLI 354 EWPFDADRWVISDFIQLPDDVTSKVKELQRVHQKRFMVAWSAMGXP6I 403
          EWPFDADRWVISDFIQLPDDVTSKVKELQRVHQKRFMVAWSAMGXP6I
gi|2367274|gb|AAC76768.1| 354 EWPFDADRWVISDFIQLPDDVTSKVKELQRVHQKRFMVAWSAMGXP6I 403

YIEV_ECOLI 404 MRIFD 408
          MRIFD
gi|2367274|gb|AAC76768.1| 404 MRIFD 408

```

If you think there is anything wrong with this script, please contact [Pfam](#)

FIG. 13

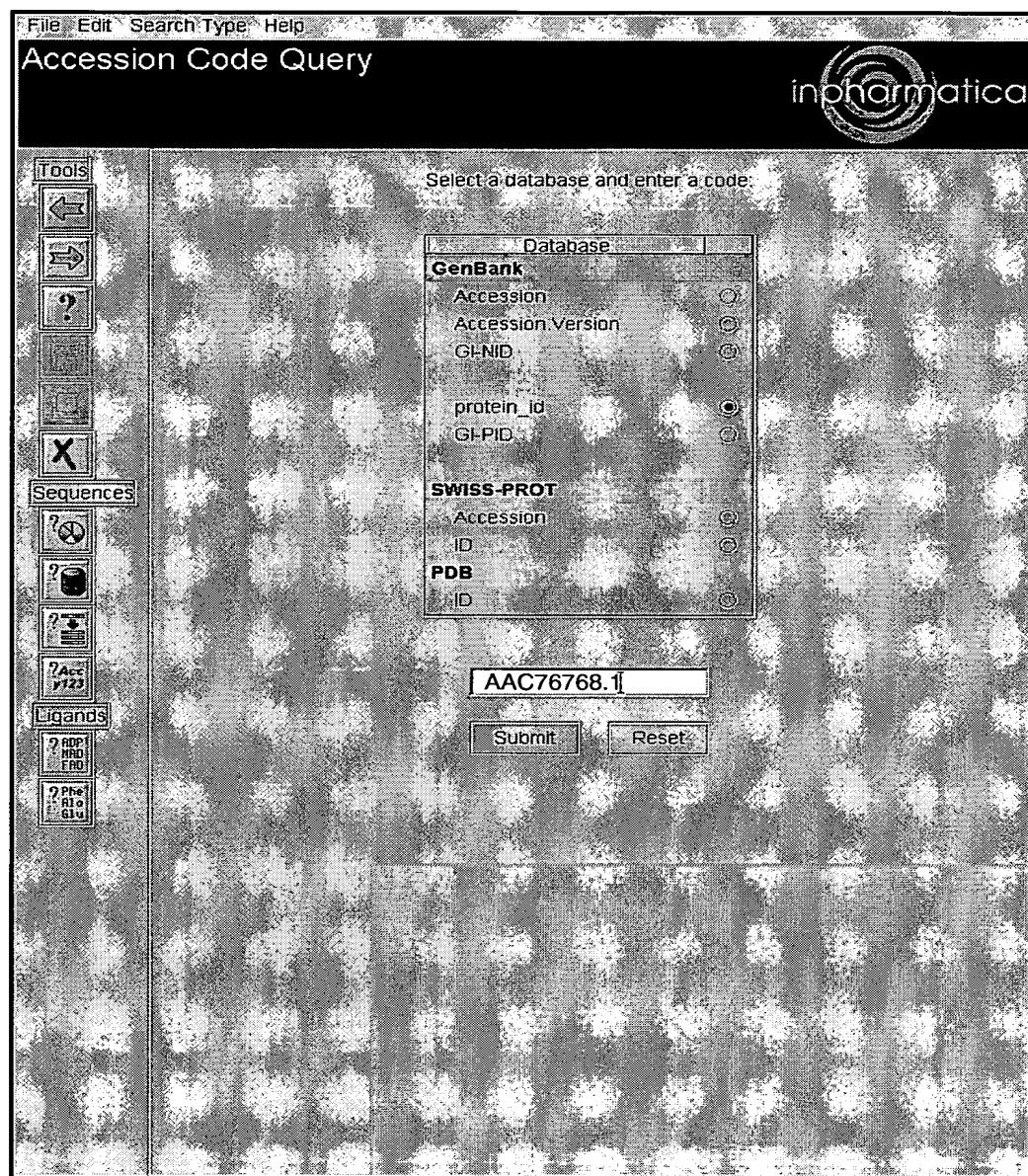
19/40

File Edit View Go Communicator Help

Bookmarks Location: <http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=protein>

LOCUS	ARC76768	427 aa	BCT	01-DEC-2000
DEFINITION	orf, hypothetical protein [Escherichia coli K12].			
ACCESSION	ARC76768			
PID	g2367274			
VERSION	ARC76768.1	GI:2367274		
DBSOURCE	locus <a href="#">RE000451</a>	accession <a href="#">RE000451.1</a>		
KEYWORDS				
SOURCE	Escherichia coli K12			
ORGANISM	Escherichia coli K12			
	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.			
REFERENCE	1 (residues 1 to 427)			
AUTHORS	Blattner, F.R., Plunkett, G. III, Bloch, C.R., Perna, N.T., Burland, B., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, M.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.			
TITLE	The complete genome sequence of Escherichia coli K-12			
JOURNAL	Science 277 (5331), 1453-1474 (1997)			
MEDLINE	97426517			
PUBMED	<a href="#">9278503</a>			
REFERENCE	2 (residues 1 to 427)			
AUTHORS	Blattner, F.R.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: <a href="mailto:ecoli@genetics.wisc.edu">ecoli@genetics.wisc.edu</a> Phone: 608-262-2534 Fax: 608-263-7459			
REFERENCE	3 (residues 1 to 427)			
AUTHORS	Blattner, F.R.			
TITLE	Direct Submission			
JOURNAL	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: <a href="mailto:ecoli@genetics.wisc.edu">ecoli@genetics.wisc.edu</a> Phone: 608-262-2534 Fax: 608-263-7459			
REFERENCE	4 (residues 1 to 427)			
AUTHORS	Plunkett, G. III.			
TITLE	Direct Submission			
JOURNAL	Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA			
COMMENT	This sequence was determined by the <i>E. coli</i> Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIMH grants M600301 and M601428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from <i>E. coli</i> K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: <a href="mailto:mark@amber.gatech.edu">mark@amber.gatech.edu</a> ]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the <i>E. coli</i> Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlin. A public version of the database is accessible ( <a href="http://cgsc.biology.yale.edu">http://cgsc.biology.yale.edu</a> ). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. comments to the authors are appreciated. Updated information will be available at the <i>E. coli</i> Genome Project's World Wide Web site ( <a href="http://www.genetics.wisc.edu">http://www.genetics.wisc.edu</a> ). *** The <i>E. coli</i> K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.			
FEATURES	Method: conceptual translation.			
source	Location/Qualifiers			
	1..427			
	/organism="Escherichia coli K12"			
	/strain="K12"			
	/sub_strain="MG1655"			
	/db_xref="taxon:83333"			
protein	1..427			
	/function="orf; Unknown"			
	/product="orf, hypothetical protein"			
CDS	1..427			
	/gene="piem"			
	/coded_by="complement(2367272..5249..6532)"			
	/transl_table=11			
	/note="f427; sequence change joins ORFs piem and piem from earlier version"			
ORIGIN	1 mrsrlkdarv ppelteewmc yqqsql1stp qfivqlppqil dlhrinspw aeqarqlvda 61 nstitsalht lfigrwrlsl ivgattlmgg lletereqlgq sevgematsls gqlepliladn 121 ntaagrlwdm sagilkrgdy qlivkypgefl neppelkrla eqlgrsreak siphndaqme 181 tfrtmvrep a typeqvvdg1q qsd6d1rlp relatloite lensfnrnu sknltnrh			

FIG. 14A



## FIG. 14B Aligned Sequence Display

Query details:  
AAC76768.1: orf, hypothetical protein

File Edit Search Type Help

Total selected 0      View alignment

Page 1 of 3      Total hits 88      Selected 0

Select all       Deselect all

Inpharmatica Genome Threader results

Cluster	Detail	Code	Type	Organism	%ID	Query	Target	Align score	Method	Confidence
•	1CK4B	CRYSTAL STRUCTURE OF RAT A1B1 INTEGRIN DOMAIN FROM INTEGRIN CH3/MG2* BO...	<i>Rattus norvegicus</i>	12.6	267 - 398	6 - 144	159	Local	Certain(100%)	
•	1LJM	1-DOMAIN FROM INTEGRIN CH3 MN2* BO...	<i>Homo sapiens</i>	19.8	257 - 394	5 - 130	157	Local	Certain(100%)	
•	1CK4A	CRYSTAL STRUCTURE OF RAT A1B1 INTEGRIN CH3 DOMAIN IN MAGNESIUM COMPLEX	<i>Rattus norvegicus</i>	12.6	267 - 398	6 - 144	149	Local	Certain(100%)	
•	1BHQ1	MAC-1 DOMAIN CADMIUM COMPLEX	<i>Homo sapiens</i>	19.8	267 - 394	5 - 130	146	Local	Certain(100%)	
•	1BHQ2	MAC-1 DOMAIN CADMIUM COMPLEX	<i>Homo sapiens</i>	19.8	267 - 394	5 - 130	146	Local	Certain(100%)	
•	1BHQ1	MAC-1 DOMAIN CADMIUM COMPLEX	<i>Homo sapiens</i>	19.8	267 - 394	5 - 130	146	Local	Certain(100%)	
•	1BHQ2	MAC-1 DOMAIN CADMIUM COMPLEX	<i>Homo sapiens</i>	19.8	267 - 394	5 - 130	146	Local	Certain(100%)	
•	1BHQ1	MAC-1 DOMAIN CADMIUM FREE	<i>Homo sapiens</i>	19.8	267 - 394	5 - 130	146	Local	Certain(100%)	
•	1BHQ1	MAC-1 DOMAIN CADMIUM FREE	<i>Homo sapiens</i>	19.8	267 - 394	5 - 130	146	Local	Certain(100%)	
•	1BHQ2	MAC-1 DOMAIN CADMIUM FREE	<i>Homo sapiens</i>	19.8	267 - 394	5 - 130	146	Local	Certain(100%)	
•	1LFB	CD1A 1-DOMAIN WITH BOUND MN++	<i>Homo sapiens</i>	15.9	264 - 381	1 - 124	143	Local	Certain(100%)	
•	1ZON	CD1A 1-DOMAIN WITHOUT BOUND CAT++	<i>Homo sapiens</i>	15.9	264 - 381	1 - 124	143	Local	Certain(100%)	
•	1LFAA	CD1A 1-DOMAIN WITH BOUND MN++	<i>Homo sapiens</i>	14.6	264 - 354	1 - 96	141	Local	Certain(100%)	
•	1ZOA	CD1A 1-DOMAIN WITH BOUND MAGNESIUM	<i>Homo sapiens</i>	14.6	264 - 354	1 - 96	141	Local	Certain(100%)	
•	1ZOPB	CD1A 1-DOMAIN WITH BOUND MAGNESIUM	<i>Homo sapiens</i>	14.6	264 - 354	1 - 96	141	Local	Certain(100%)	
•	1ZOPA	CD1A 1-DOMAIN WITH BOUND MAGNESIUM	<i>Homo sapiens</i>	14.6	264 - 354	1 - 96	141	Local	Certain(100%)	
•	1ZOOB	CD1A 1-DOMAIN WITH BOUND MAGNESIUM	<i>Homo sapiens</i>	14.6	264 - 354	1 - 96	141	Local	Certain(100%)	
•	1DGG4	IN VS SOLUTION STRUCTURE OF THE INSERT	<i>Homo sapiens</i>	14.3	267 - 354	10 - 100	139	Local	Certain(100%)	
•	1ATZA	HUMAN VON WILLEBRAND FACTOR A3 DOMAIN	<i>Homo sapiens</i>	11.3	267 - 407	6 - 152	135	Local	Certain(100%)	
•	1OXA	CRYSTAL STRUCTURE OF THE VON WILLEBRAND FACTOR A3 DOMAIN	<i>Homo sapiens</i>	11.3	267 - 393	11 - 150	111	Local	Certain(100%)	
→	1AOXA	DOMAIN FROM INTEGRIN ALPHA2-BETA1	<i>Homo sapiens</i>	13.2	267 - 384	9 - 134	107	Local	Certain(100%)	
•	1QCSB	DOMAIN FROM INTEGRIN ALPHA1-BETA1	<i>Homo sapiens</i>	11.9	267 - 398	5 - 133	110	Local	Certain(100%)	
•	1QCSA	DOMAIN FROM INTEGRIN ALPHA1-BETA1	N/A	11.9	267 - 398	7 - 145	102	Local	Certain(100%)	
•	1AOXB	DOMAIN FROM INTEGRIN ALPHA2-BETA1	<i>Homo sapiens</i>	9.8	267 - 398	1 - 149	98	Local	Certain(100%)	
•	1AOXA	AS DOMAIN OF VON WILLEBRAND FACTOR A3 DOMAIN	<i>Homo sapiens</i>	10.7	267 - 398	8 - 145	98	Local	Certain(100%)	
•	1AOXB	AS DOMAIN OF VON WILLEBRAND FACTOR A3 DOMAIN	<i>Homo sapiens</i>	10.7	267 - 398	8 - 145	98	Local	Certain(100%)	
•	1AUG	AS DOMAIN OF VON WILLEBRAND FACTOR A3 DOMAIN	<i>Homo sapiens</i>	11.8	250 - 398	4 - 159	97	Local	Certain(100%)	
•	1AIZB	HUMAN VON WILLEBRAND FACTOR A3 DOMAIN	<i>Homo sapiens</i>	10.7	267 - 398	9 - 146	96	Local	Certain(100%)	
•	1AJJB	NITROGEN REGULATORY BACTERIAL PROTEIN	<i>Escherichia coli</i>	18.6	306 - 348	13 - 57	60	Local	Low(42%)	
•	1BKDS	COMPLEX OF HUMAN H-4 PASSIVE	<i>Homo sapiens</i>	14.4	237 - 302	69 - 202	69	Local	Low(34%)	

GenBank SWISS-PROT PDB

# FIG. 14C

## Aligned Sequence Display

Query details:  
AAC76768.1: orf, hypothetical protein

File Edit Search Type Help

Total selected: 0 View alignment

Reverse Maximised PSI-BLAST results

Page: 1 of 17 Total hits: 508 Selected: 0 Select all Deselect all

Cluster	Detail	Code	Title	Organism	%ID	Query res.	Target res.	Bliter	E-value
o	►	BAA80016.1	452aa long hypothetical protein	<i>Artemia</i>	28.0	156 - 407	202 - 431	2	2.00E-60
o	►	Q3D394	HYPOTHETICAL PROTEIN [Mj0077]	<i>M. marinascifli</i>	22.0	146 - 408	122 - 361	3	2.00E-71
o	►	AAB88810.1	conserved hypothetical protein	<i>M. marinascifli</i>	23.0	147 - 410	170 - 420	3	2.00E-65
o	►	AAC75810.1	putative retroelement polypeptide	<i>Arthroleptis thalassinus</i>	19.0	168 - 404	124 - 403	3	2.00E-50
o	►	Q55874	HYPOTHETICAL 48 kDa PROTEIN [SL0103]	<i>S. sp.</i>	21.0	241 - 407	17 - 188	3	4.00E-39
o	►	AAB80957.1	Magnesium-chelatase subunit [Chi]	<i>Arthroleptis thalassinus</i>	25.0	15 - 370	324 - 535	3	5.00E-22
o	►	CAA30478.1	Not given	<i>M. musculus</i>	20.0	247 - 416	132 - 309	4	2.00E-44
o	►	AAA58544.1	Not given	<i>H. sapiens</i>	115.0	223 - 412	107 - 305	4	3.00E-42
o	►	AAB24821.1	leukocyte integrin alpha chain	<i>H. sapiens</i>	16.0	223 - 412	107 - 305	4	3.00E-42
o	►	1JNM	L-DOMAIN FROM INTEGRIN CR3, MN2 - BOUND	<i>H. sapiens</i>	15.0	268 - 412	1 - 158	4	2.00E-41
o	►	1ID0	L-DOMAIN FROM INTEGRIN CR3, MG2 - BOUND	<i>H. sapiens</i>	16.0	263 - 412	1 - 158	4	2.00E-41
o	►	1HQ1	MAC-1 DOMAIN MAGNESIUM COMPLEX	<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	►	1IDN2	MAC-1 DOMAIN METAL FREE	<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	►	1IDN1	MAC-1 DOMAIN METAL FREE	<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	►	1HQ02	MAC-1 DOMAIN MAGNESIUM COMPLEX	<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	►	1HQ02	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	►	1HQ01	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	►	Q22437	MAGNESIUM-CHELATASE SUBUNIT CHILD PRECURSOR	<i>P. sativum</i>	19.0	162 - 412	358 - 701	4	7.00E-41
o	►	CAB69668.1	Not given	<i>Unidentified</i>	18.0	162 - 417	473 - 705	4	3.00E-41
o	►	QAB16565.1	CD11b	<i>S. cerevisiae</i>	17.0	254 - 417	6 - 177	4	8.00E-40
o	►	CAB58178.1	Mg-protoporphyrin IX	<i>H. vulgare</i>	17.0	162 - 410	90 - 311	4	2.00E-38
o	►	QAD52031.1	magnesium-chelatase subunit CHILD precursor	<i>A. thaliana</i>	17.0	162 - 423	441 - 695	4	5.00E-38
o	►	BAA16787.1	Mg chelatase subunit	<i>S. sp.</i>	18.0	188 - 367	415 - 588	4	6.00E-38
o	►	AAB88581.1	conserved protein	<i>M. marinascifli</i>	13.0	154 - 383	60 - 309	4	7.00E-37
o	►	CAA77537.1	661 aa (60 kD) Mg chelatase subunit	<i>R. capsulatus</i>	21.0	168 - 367	277 - 484	4	1.00E-36
o	►	QAF22895.1	127G7.20	<i>A. thaliana</i>	17.0	152 - 223	491 - 747	4	2.00E-36
o	►	AAF21241.1	alpha D Integrin	<i>Annona squamosa</i>	14.0	238 - 412	126 - 307	4	2.00E-35
o	►	AAF23892.1	leukocyte adhesion glycoprotein isoform alpha/beta/gamma	<i>M. marinascifli</i>	13.0	231 - 412	118 - 307	4	2.00E-35
o	►	AAB88547.1	leukocyte adhesion alpha chain	<i>H. sapiens</i>	13.0	239 - 410	124 - 303	4	3.00E-35

GenBank SWISS-PROT D2DB

FIG. 15

AIEye output (December 13, 2000 3:07 PM)

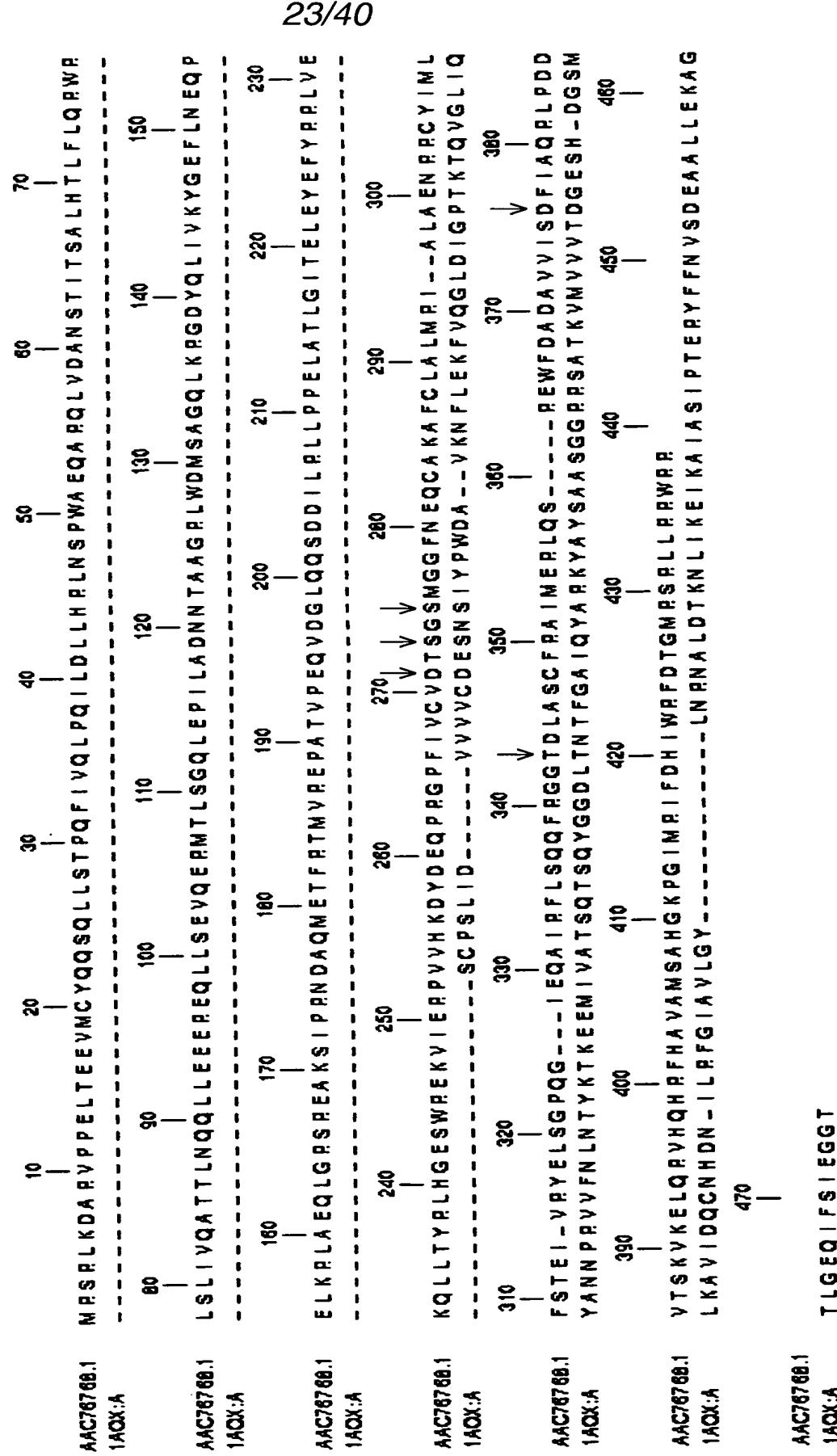


FIG. 16B

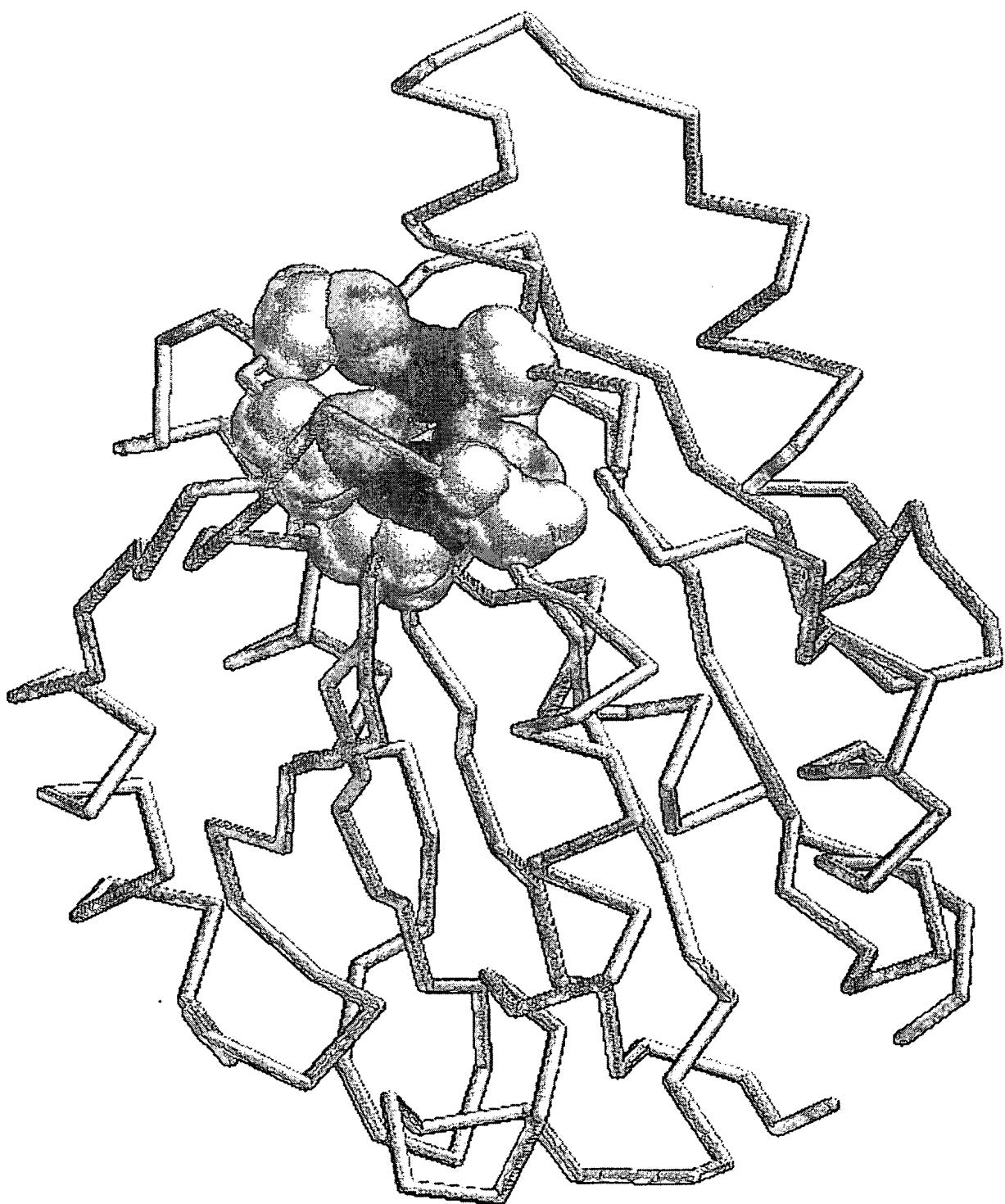
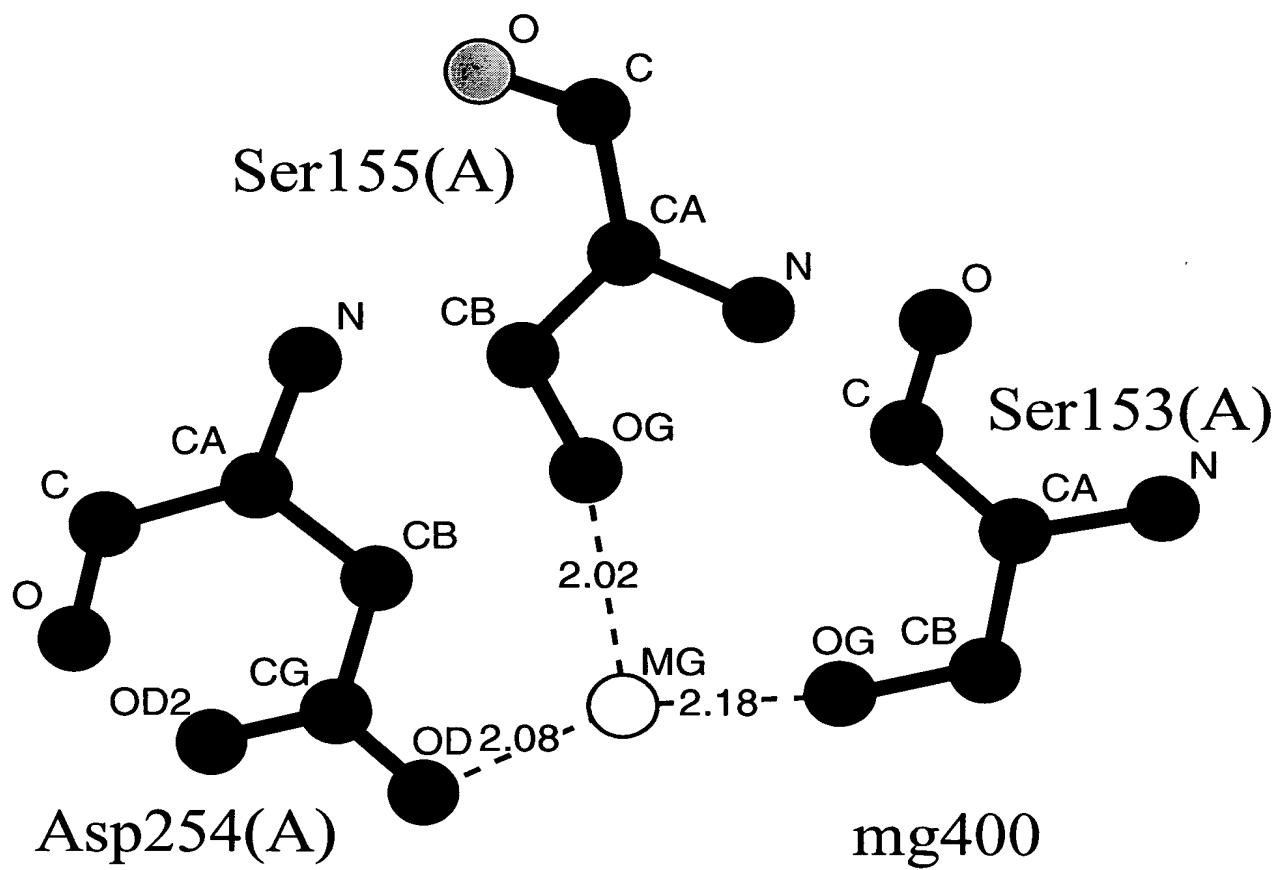


FIG. 16A



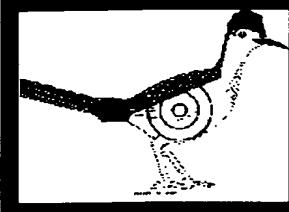
1aox: MG400

FIG. 17

26/40

File Edit View Go Communicator Help

Bookmarks Location http://victoria.inpharmatica.co.uk/~volker/BPD3target.html

 Target Mining Interface



**Select Your Query Sequence**

• Enter PDB accession number (e.g. 1QMA):  and chain (e.g. B):

OR

• Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

**Select Database**

Release:

**Apply Filters**

• Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first\_PB\_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

• Filter for the following SPECIES:

Homo sapiens  Rattus norvegicus (Rat)  Mus musculus (Mouse)  Danio rerio (Zebra fish)

100%

FIG. 18A

27/40

File Edit View GO Communicator Help

Forward Related Home Search Netscape Print Security Stop

Bookmarks Location: http://London-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl

2) 81 additional hits identified by both, Genome Threader and PSI-BLAST:

Combined Genome Threader and PSI - Blast output: PSI - BLAST values are shown in maroon!

Add2list	BPD links	WWW link	Title	Organism	Div.	%ID (GT,PSI)	Query rgn. (GT,PSI)	Target rgn. (GT,PSI)	Aln. score (GT)	Conf. (GT)	1st Iter. (PSI)
↓	<a href="#">AAA59544.1</a> drill through Top50BlastHits	<a href="#">AAA59544.1</a>	Not given	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	100% unmaskedGT	1
↓	<a href="#">AAB24821.1</a> drill through Top50BlastHits	<a href="#">AAB24821.1</a>	leukocyte integrin alpha chain	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	100% unmaskedGT	1
↓	<a href="#">Q99715</a> drill through Top50BlastHits	<a href="#">Q99715</a>	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human)	PRI	28.9%, 28% unmaskedSW	2-186, 2-179	439-617, 2322-2494	456	100% unmaskedGT	1
↓	<a href="#">AAB38702.1</a> drill through Top50BlastHits	<a href="#">AAB38702.1</a>	cartilage matrix protein	Homo sapiens	PRI	28.9%, 25% unmaskedSW	2-186, 2-186	274-452, 40-221	446	100% unmaskedGT	1
↓	<a href="#">AAC01506.1</a> drill through Top50BlastHits	<a href="#">AAC01506.1</a>	type XII collagen	Homo sapiens	PRI	28.4%, 28% unmaskedSW	2-186, 2-186	137-318, 137-318	445	100% unmaskedGT	1
↓	<a href="#">CAA72402.1</a> drill through Top50BlastHits	<a href="#">CAA72402.1</a>	collagen type XIV	Homo sapiens	PRI	28.7%, 30% unmaskedSW	2-186, 2-186	6-185, 6-185	442	100% unmaskedGT	1
↓	<a href="#">AAB38547.1</a> drill through Top50BlastHits	<a href="#">AAB38547.1</a>	leukointegrin alpha d chain	Homo sapiens	PRI	61%, 60% unmaskedSW	1-187, 1-187	148-334, 148-334	439	100% unmaskedGT	1
↓	<a href="#">CAB71222.1</a> drill through Top50BlastHits	<a href="#">CAB71222.1</a>	dJ238D15.1 (collagen, type XII, alpha 1)	Homo sapiens	PRI	27.1%, 22% unmaskedSW	1-186, 2-186	293-472, 1430-1620	439	100% unmaskedGT	1
↓	<a href="#">CAA07569.1</a> drill through Top50BlastHits	<a href="#">CAA07569.1</a>	matrilin-4	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-187	344-525, 34-215	418	100% unmaskedGT	2
↓	<a href="#">CAB46380.1</a> drill through Top50BlastHits	<a href="#">CAB46380.1</a>	dJ453C12.3 (matrilin-4)	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-186	385-566, 34-214	418	100% unmaskedGT	2

FIG. 18B

28/40

KIAA0835 protein									
-	Bed Sea View	ctrl through Tos50BlastH1s	BA01611..						
-	Bed Sea View	ctrl through Tos50BlastH1s	CA052192..1	CA052192..1	G7c protein				
-	Bed Sea View	ctrl through Tos50BlastH1s	BA075895..1	BA075895..1	N-copine				
-	Bed Sea View	ctrl through Tos50BlastH1s	BA075899..1						
-	Bed Sea View	ctrl through Tos50BlastH1s	PA0012..	PA0012..	INTEGRIN BETA-8 PRECURSOR.				
-	Bed Sea View	ctrl through Tos50BlastH1s	AA0211820..1	AA0211820..1					
-	Bed Sea View	ctrl through Tos50BlastH1s	CA065775..1	CA065775..1	Sec23 protein				
-	Bed Sea View	ctrl through Tos50BlastH1s	AA015920..1	AA015920..1	copine I				
-	Bed Sea View	ctrl through Tos50BlastH1s	AA015920..1	AA015920..1					
-	Bed Sea View	ctrl through Tos50BlastH1s	AA015920..1	AA015920..1	polymerase				
-	Bed Sea View	ctrl through Tos50BlastH1s	CA0897610..1	CA0897610..1	dJ039K20..3 (Copine I (similar to KIA0836))				
-	Bed Sea View	ctrl through Tos50BlastH1s	AA015532..1	AA015532..1	Not I protein				
-	Bed Sea View	ctrl through Tos50BlastH1s	PA0155..	PA0155..	60 kDa R PROTEIN (60 kDa RIBONUCLEOPROTEIN PO)(RNP)(SJOGREN SYNDROME TYPE A ANTIGEN (SS-A))				
-	Bed Sea View	ctrl through Tos50BlastH1s	AA015683..1	AA015683..1	hexokinase I				
-	Bed Sea View	ctrl through Tos50BlastH1s	BA092672..1	BA092672..1	KIAA1434 protein				
-	Bed Sea View	ctrl through Tos50BlastH1s	CAA0035..1	CAA0035..1	Sec24B protein				
-	Bed Sea View	ctrl through Tos50BlastH1s	CA010235..1	CA010235..1					
-	Bed Sea View	ctrl through Tos50BlastH1s	AA013765..1	AA013765..1	protoactinin gamma A11 short form protein				
-	Bed Sea View	ctrl through Tos50BlastH1s	AA013766..1	AA013766..1					
-	Bed Sea View	ctrl through Tos50BlastH1s	AA013714..1	AA013714..1	protoactinin gamma A11				
-	Bed Sea View	ctrl through Tos50BlastH1s	CA008933..1	CA008933..1	annexin 31 (annexin XXXI)				
-	Bed Sea View	ctrl through Tos50BlastH1s	AA059165..1	AA059165..1	integrin beta-7 subunit				
-	Bed Sea View	ctrl through Tos50BlastH1s	AA065542..1	AA065542..1	integrin variant beta4E				
-	Bed Sea View	ctrl through Tos50BlastH1s							

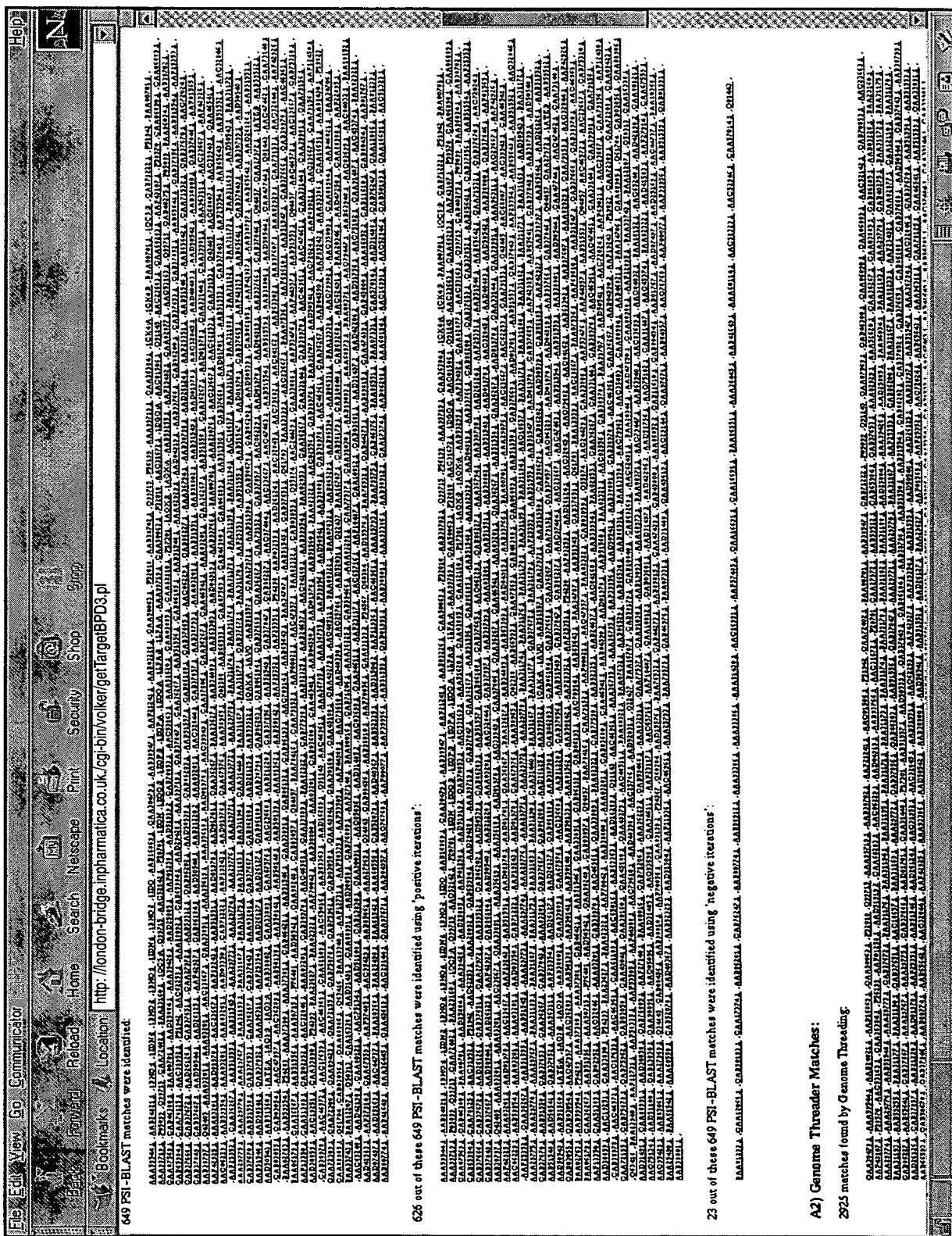


FIG. 18C

Biopendium additional annotation page - Netscape

File Edit View Go Communicator Help

Back Forward Home Search Netscape Print Help

Bookmarks Location: /i/london-bridge/inpharmatica/BPDDEV7/cgi-bin/spl?/ep.esn=55817&password=cats\_appusername=cats\_app&oracle\_sid=BPDDEV7 What's Related

Aligned annotation view for P10155 (downloading image...)

Primary database information:  Swiss-Prot

Secondary database information:  Prosite matches

Inpharmatica calculated information:  Protein matches  Ligand binding residues  Masked residues

1 50 100 150 200 250 300 350 400 450 500

1: AA03550.1  
Rep: P10155

Request information

Source database: SWISSPROT

Accession code: P10155

Gene name: SSA2 OR R060

Download sequence in FASTA format

Links to other resources:

View custom hyperlinks

Search in ExPASy

European Bioinformatics Institute EBI000504

Online Mendelian Inheritance in Man 234700

Online Mendelian Inheritance in Man 600063

Protein Information Resource A30596

Protein Information Resource A31760

Swissprot comments:

FUNCTION UNKNOWN

FO SMALL RIBONUCLEOPROTEINS CONSIST OF FOUR SMALL RNA MOLECULES OF 85-112 NT EACH OF

Document Done

FIG. 19

FIG. 20

File Edit View Go Communicator Help

Blank Forward Reload Home Search Netscape Print Security Stop

Bookmarks Location: http://www.sanger.ac.uk/cgi-bin/Pfam/nph-search.cgi

**The Sanger Centre** **Pfam**  
Protein families database of alignments and HMMs

Home | Keyword search | Protein search | DNA search | Browse Pfam | Taxonomy search | Help

**Results for gi|133251|sp|P10155|RO60\_HUMAN**

There were no matches to Pfam-A (including borderline matches) for gi|133251|sp|P10155|RO60\_HUMAN

**Matches to Pfam-B**

Domain	Start	End	Evalue	Alignment
Pfam-B_8344	1	194	2.3e-103	Align
Pfam-B_10162	195	538	1.8e-165	Align

[538 residues]

**Alignments of Pfam-B domains to best-matching Pfam-B sequence**

Format for fetching alignments to Pfam-B families: Hypertext linked to swisspfam

Query gi|133251|sp|P10155|RO60\_HUMAN/1-194 matching Pfam-B\_8344

```

Q92787 1 MEESVNQMQPLNEKQIANSQDGYYVWQVTDMNRLKQRFPLCFG6SEGTYYYIKE 50
       1 MEESVNQMQPLNEKQIANSQDGYYVWQVTDMNRLKQRFPLCFG6SEGTYYYIKE 50
gi|133251|sp|P10155|RO60_HUMAN 1 MEESVNQMQPLNEKQIANSQDGYYVWQVTDMNRLKQRFPLCFG6SEGTYYYIKE 50
Q92787 51 QKLGLENAEALIRLIEDGRGCEVIQEIKSFSQEGRTTKQEPMLFALACCS 100
       QKLGLENAEALIRLIEDGRGCEVIQEIKSFSQEGRTTKQEPMLFALACCS 100
gi|133251|sp|P10155|RO60_HUMAN 51 QKLGLENAEALIRLIEDGRGCEVIQEIKSFSQEGRTTKQEPMLFALACCS 100
Q92787 101 QCSDISTKQARAFKAVSEVCRIPTHLPTFIQFKKDLKESMKCGMWGRALRK 150
       QCSDISTKQARAFKAVSEVCRIPTHLPTFIQFKKDLKESMKCGMWGRALRK 150
gi|133251|sp|P10155|RO60_HUMAN 101 QCSDISTKQARAFKAVSEVCRIPTHLPTFIQFKKDLKESMKCGMWGRALRK 150
Q92787 151 RIAWDWYNEKGGMALALAUTVKYKQRNGWSMKDLLRLSMLKPSSEG 194
       RIAWDWYNEKGGMALALAUTVKYKQRNGWSMKDLLRLSMLKPSSEG 194
gi|133251|sp|P10155|RO60_HUMAN 151 RIAWDWYNEKGGMALALAUTVKYKQRNGWSMKDLLRLSMLKPSSEG 194

```

**Align to family**

Query gi|133251|sp|P10155|RO60\_HUMAN/195-538 matching Pfam-B\_10162

```

008848 195 LAIVTKYITKGWKEVHEEYKEKALSVEAKLLKYLEAVEVKVKRTKDDLEV 244
       LAIVTKYITKGWKEVHEEYKEKALSVEAKLLKYLEAVEVKVKRTKDDLEV 244
gi|133251|sp|P10155|RO60_HUMAN 195 LAIVTKYITKGWKEVHEEYKEKALSVEAKLLKYLEAVEVKVKRTKDDLEV 244
008848 245 IMLIEENQLVREMLILTHMLKSKEVVKALLQEMPLTALLRLNLGKMTANSVL 294
       IMLIEENQLVREMLILTHMLKSKEVVKALLQEMPLTALLRLNLGKMTANSVL
008848 295 EPGHSEVSLICEKLNSNEKLLKKARIMPMPHILIALETYPRGNGLRLGKLW 344
       EPGHSEVSL+CEKLNEKLLKKARIMPMPH+IALETY+ GMGLRGKLW
gi|133251|sp|P10155|RO60_HUMAN 295 EPGHSEVSLVCEKLNSNEKLLKKARIMPMPHILIALETYPRGNGLRLGKLW 344
008848 345 PDKDILQALDRAYTTFKTVEPTGKRFLLRVDVSASMNQRALGSVLNAST 394

```

## FIG. 21

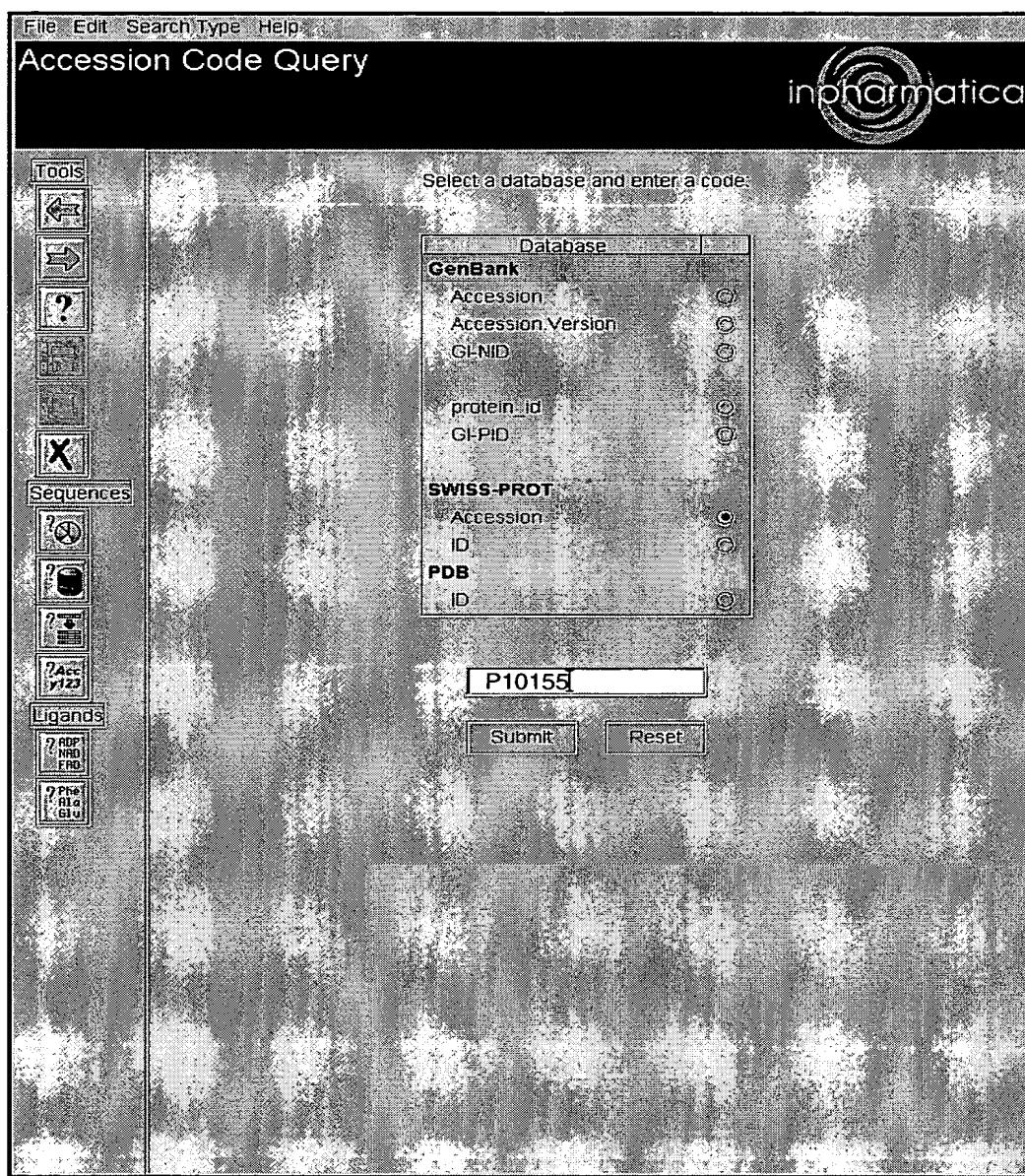
File Edit View Go Communicator Help

Bookmarks Location: <http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=Prot>

**LOCUS** R060\_HUMAN 538 aa **PRI** 01-FEB-1996  
**DEFINITION** 60 KD RO PROTEIN (60 KD RIBONUCLEOPROTEIN RO) (R0RNP) (SJOGREN SYNDROME TYPE A ANTIGEN (SS-A)).

**ACCESSION** P10155  
**PID** g133251  
**VERSION** P10155 GI:133251  
**DBSOURCE** swissprot: locus R060\_HUMAN, accession [P10155](#); class: standard.  
 created: Mar 1, 1989.  
 sequence updated: Mar 1, 1989.  
 annotation updated: Feb 1, 1996.  
 xrefs: gi: gi: [177782](#) gi: gi: [177783](#), gi: gi: [387656](#), gi: gi: [387657](#), gi: gi: [86722](#), gi: gi: [107626](#)  
 xrefs (non-sequence databases): MIM [600063](#), MIM [234700](#), PROSITE PS00030  
**KEYWORDS** Ribonucleoprotein; RNA-binding; Systemic lupus erythematosus; Antigen.  
**SOURCE** human.  
**ORGANISM** [Homo sapiens](#)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (residues 1 to 538)  
 Deutscher, S.L., Harley, J.B. and Keene, J.D.  
**AUTHORS** Molecular analysis of the 60-kDa human Ro ribonucleoprotein  
**TITLE** Proc. Natl. Acad. Sci. U.S.A. 85 (24), 9479-9483 (1988)  
**JOURNAL** 89071722  
**MEDLINE**  
**REMARK** SEQUENCE FROM N.R.  
 2 (residues 1 to 538)  
 Ben-Chetrit, E., Gandy, B.J., Tan, E.M. and Sullivan, K.F.  
**AUTHORS** Isolation and characterization of a cDNA clone encoding the 60-kD component of the human SS-A/Ro ribonucleoprotein autoantigen  
**TITLE** J. Clin. Invest. 83 (4), 1284-1292 (1989)  
**JOURNAL** 89198084  
**MEDLINE**  
**REMARK** SEQUENCE FROM N.R.  
**COMMENT**  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. The original entry is available from <http://www.expasy.ch/sprot> and <http://www.ebi.ac.uk/sprot>  
 -----
 [FUNCTION] UNKNOWN.  
 [SUBUNIT] RO SMALL RIBONUCLEOPROTEINS CONSIST OF FOUR SMALL RNA MOLECULES OF 85-112 NT, EACH OF WHICH IS COMPLEXED WITH A 60 KD PROTEIN. RO RNPs MAY ALSO CONTAIN AN ADDITIONAL 52 KD PROTEIN.  
 [SUBCELLULAR LOCATION] CYTOPLASMIC.  
 [DISEASE] SERA FROM PATIENTS WITH SYSTEMIC LUPUS ERYTHEMATOSUS OFTEN CONTAIN ANTIBODIES THAT REACT WITH THE NORMAL CELLULAR RO PROTEIN AS IF THESE ANTIGENS WAS FOREIGN.  
 [SIMILARITY] CONTAINS 1 RNA RECOGNITION MOTIF (RNP).  
 [SIMILARITY] STRONG, TO *XENOPUS* 60 KD RO PROTEIN.  
**FEATURES**  
**source** Location/Qualifiers  
 1..538  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..538  
**Protein** 1..538  
 /product="60 KD RO PROTEIN"  
**Region** 93..98  
 /region\_name="Domain"  
 /note="RNA-BINDING (RNP2) (BY SIMILARITY)."  
**Region** 124..131  
 /region\_name="Domain"  
 /note="RNA-BINDING (RNP1) (BY SIMILARITY)."  
**Region** 239  
 /region\_name="Conflict"  
 /note="K -> R (IN REF. 2)."  
**Region** 515..538  
 /region\_name="Conflict"  
 /note="GMLDMCGFDTGALDVIRNPTLDMI -> ALQNTLLNKSP (IN REF. 2)."  
**ORIGIN**  
 1 meesvnqmqp lnekqiansq dgwwqvtdm nrlihrflcfg seggtyyike qklglenea  
 61 lirliedgrg cevigeiksf sqegrttkqe pmlfalaics qcisdistkga afkavsevcr  
 121 ipthlftfqk fkkdlkesmk cgmwgralrk aiadwpynekg gmalalavtk ykqrngwshk  
 181 dlrlshlkp sseglaivtk yitkgwkevh elykekalsv etekllkyle avekvkrtdk  
 241 elevihliee hrlvrehllt nhlkskewkk allqemplta llmlnlgkmta nsylepmse  
 301 vslvcekln eklkkarikh pfhilialet yktghglrgk lkwrpdeeil kaldaafykt  
 361 fktveptgkr fllavdvsas mnqrvlgsil nastvaaamc mvtvtrtekd s yvvaafsdemv  
 421 pcpvttdmrlt qgvlmamsqi paggttdslp miwagktntp advifivftdn etfaggvhpa  
 481 ialreyrkkm dipakliveg mtsngftiad pddrgmldmc gfdtgaldvi rnftldni  
 //

## FIG. 22A



**FIG. 22B** Aligned Sequence Display

### Query details:

P10155: 60 KDA RO PROTEIN (60 KDA RIBONUCLEOPROTEIN RO) (RORN) (SJORGREN SYNDROME TYPE A ANTIGEN (SS-A)).

# FIG. 22B Aligned Sequence Display

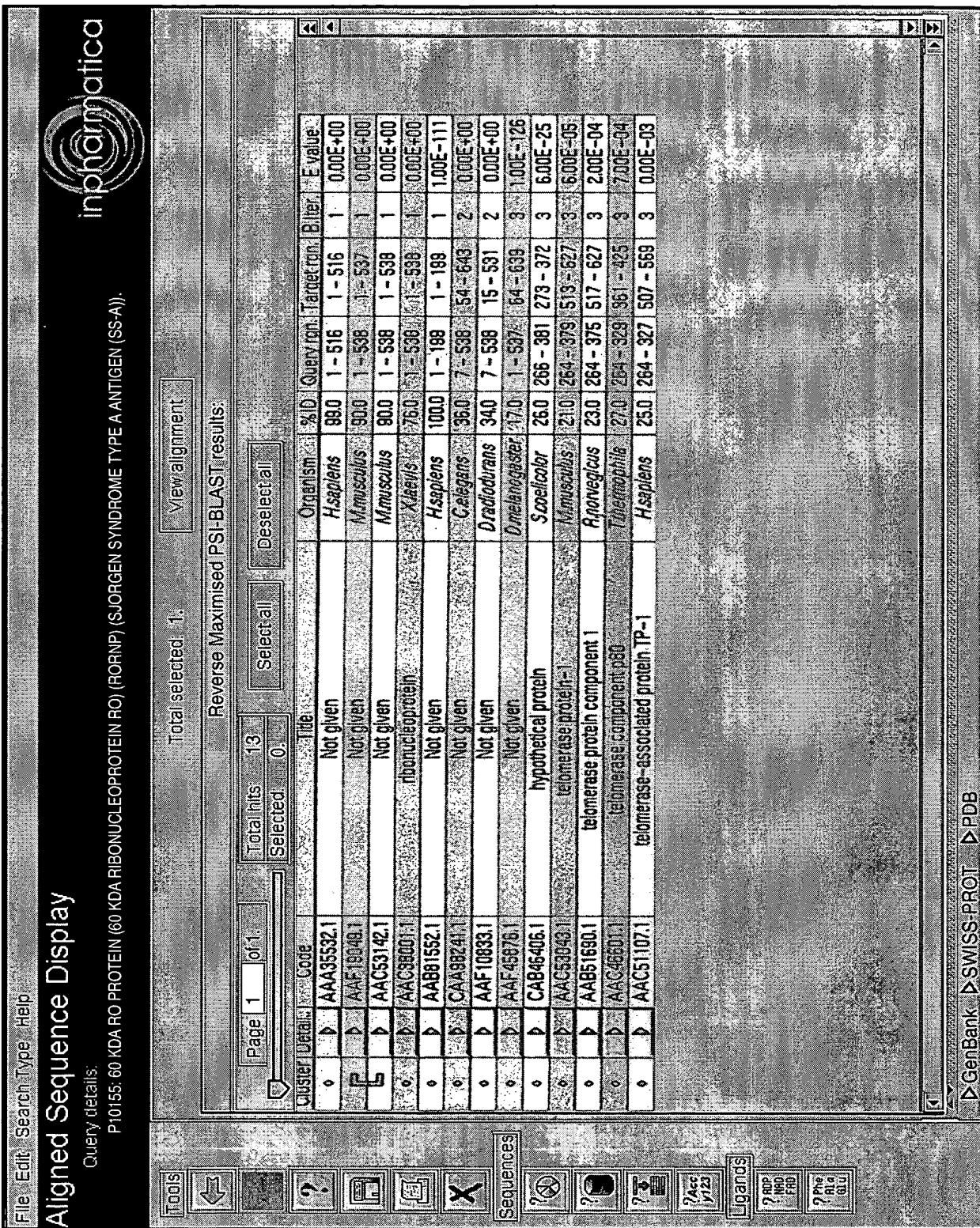


FIG. 23

AIEye output (January 2, 2001 1:29 PM)

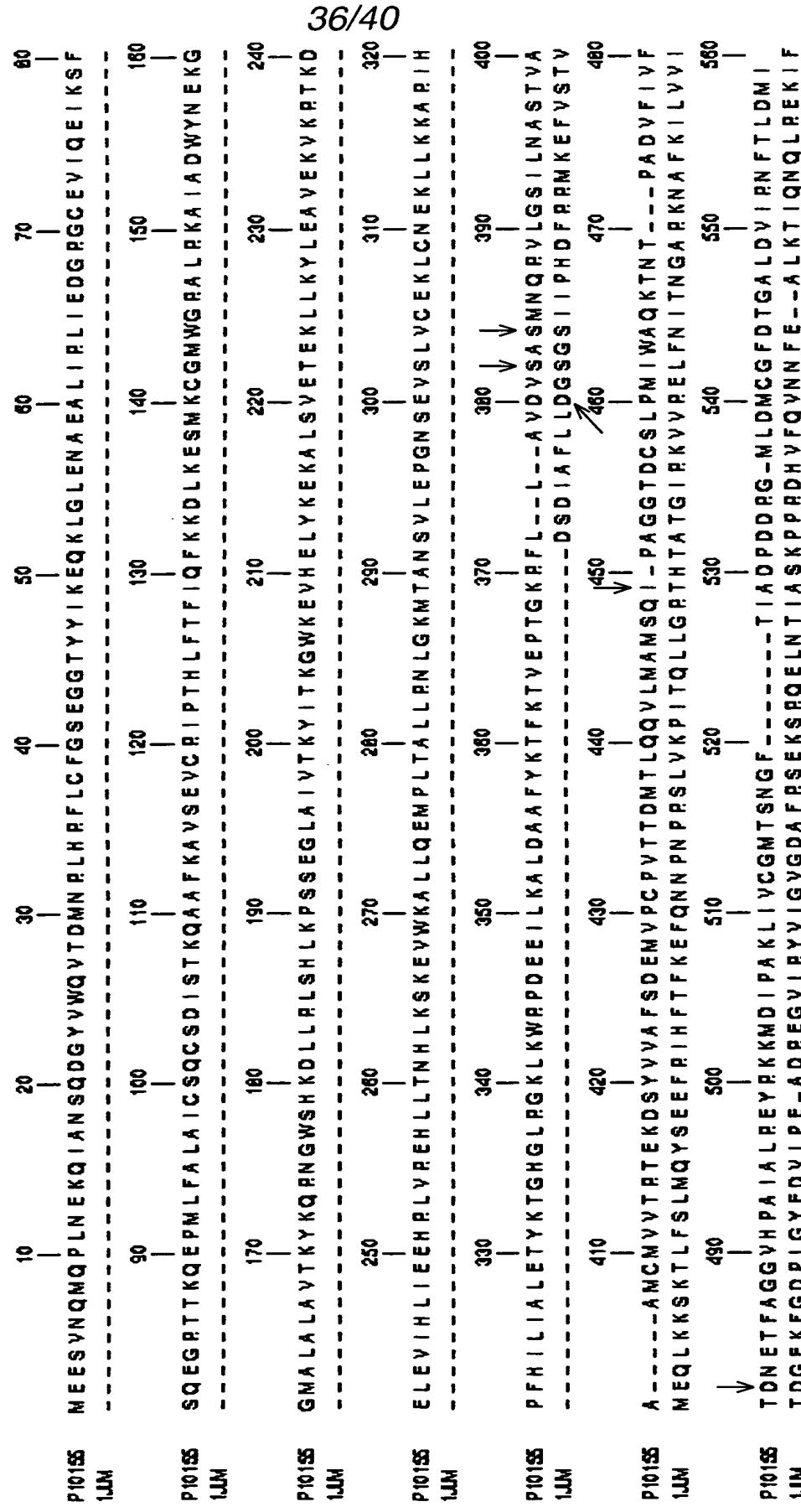


FIG. 24A

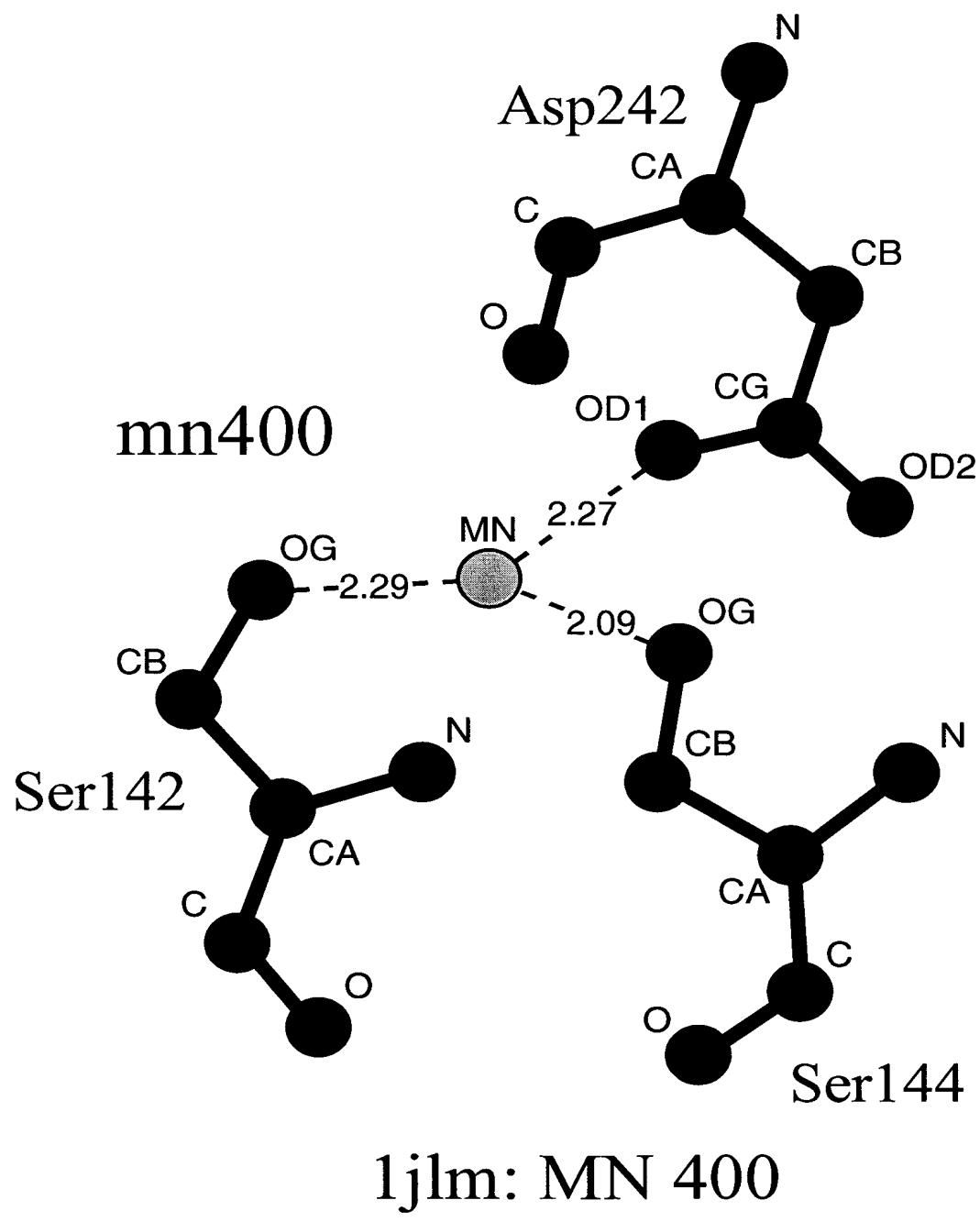


FIG. 24B

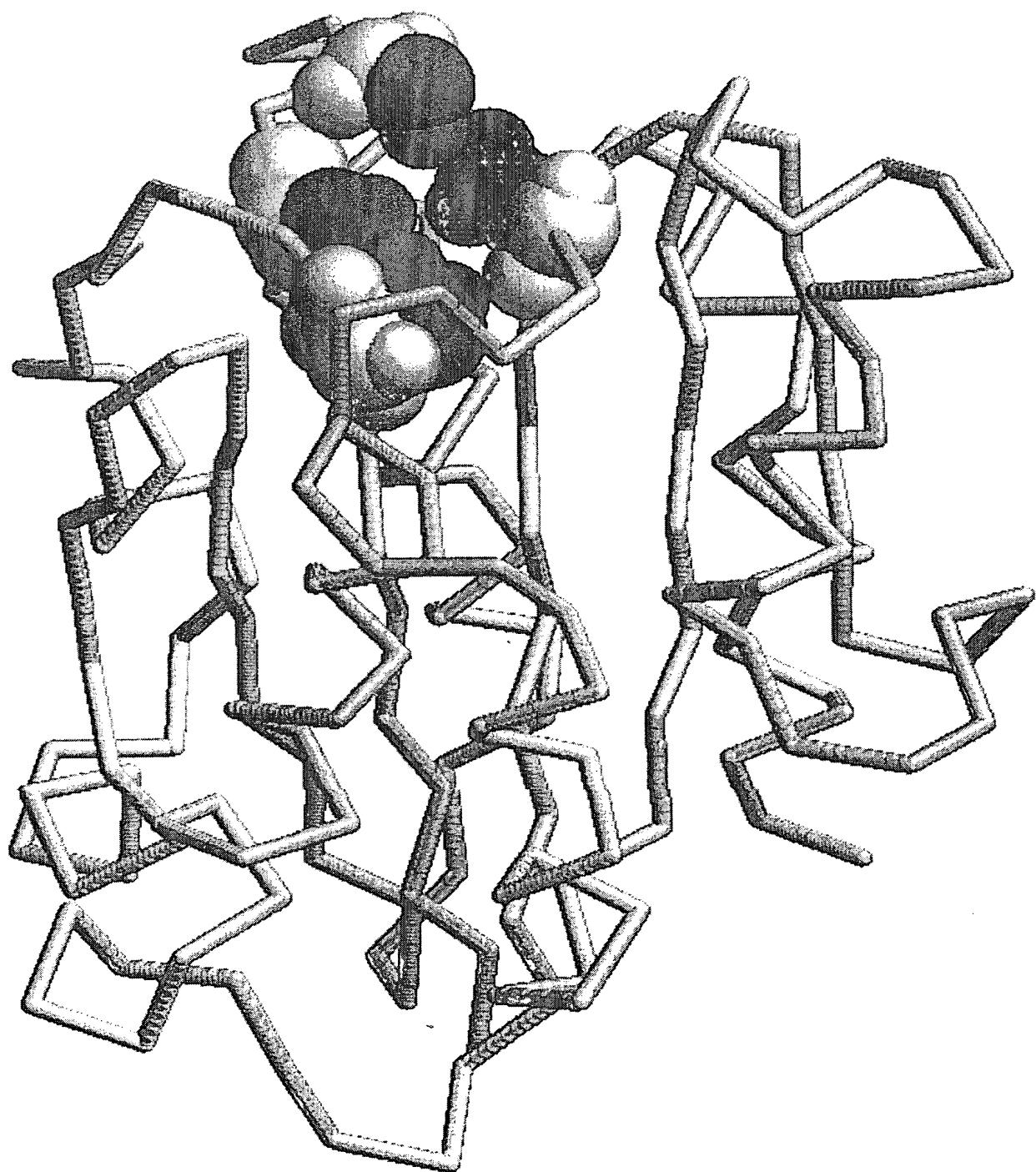


FIG. 25

AIEye output (January 4, 2001 3:18 PM)

39/40

	10	20	30	40	50	60
P10155 AAF19049.1 AAC38001.1 CAA98241.1	M E S V N - - - - - M E G S A N - - - - - M E A T M D - - - - - - - - - - M A D E L N E F Q E A G N F N E A L M R L S N V C A R L R R M Q M L E S D V E I T V V D G E L K R V P P R Q M E K V K D G Q V E N N A G G F V F P	100	110	120	130	140
P10155 AAF19049.1 AAC38001.1 CAA98241.1	V T D M N R L H R F L C F G S E G G T Y Y I K E Q K K L G L E N A E A L I R L I E D G R G C E V I Q E I K S F S Q E G R T T K Q E P M L F A L A I C S Q C S D V T D M N - V R R F L C F G S E G G T Y Y I K E Q K K L G L E N A E A L I R L I E D G R G C E V I Q E I K S F S Q E G R T A K Q E P L L F A L A V C S Q C A D V S D M N R L R R F L C F G S E G G T Y Y I E E K K L G Q E N A E A L L R L I E D G K G C E V V Q E I K T F S Q E G R A A K Q E P T L F A L A V C S Q C S D V S D E T Q V R A F L I L G S D K G S Y H Q S S E K I T I D N A Q R I I K I I E Q Q N G H M V L K E L A L I N A E N R N P K M N A M I F T L A I C A R I S T	80	90	100	110	120
P10155 AAF19049.1 AAC38001.1 CAA98241.1	160	170	180	190	200	210
P10155 AAF19049.1 AAC38001.1 CAA98241.1	I S T K Q - - - - - I N T K Q - - - - - I K T K Q - - - - - H D T T K K T E C P M L N A Y S D Y I R A L H D S A L D L I	220	230	240	250	260
P10155 AAF19049.1 AAC38001.1 CAA98241.1	E K G G M A L A L A V T K Y K Q R N G W S H K D L L R L S H - - - - - E K G G M A V A L V V T K Y K Q R N G W S H K D L L R L S H - - - - - T K D A L N L A M A V T K Y K Q R N G W S H K D L L R L S H - - - - - T K T T E K L A M L L T K Y P Q R E G G W S H R D L F R L A H P N L M D S R S H G Q S E D R L E R E Q L F R F A V K G D L V K R K R K M S V E E V A E V E K V	270	280	290	290	300
P10155 AAF19049.1 AAC38001.1 CAA98241.1	320	330	340	350	360	370
P10155 AAF19049.1 AAC38001.1 CAA98241.1	400	410	420	430	440	450
P10155 AAF19049.1 AAC38001.1 CAA98241.1	460	470	480	490	500	510

FIG. 25 (contd.)  
AIEye output (January 4, 2001 3:18 PM)

40/40

P10155 Y KTFKTV EPTGKRFLLA DVSASMMNQRVLGSILNASTVAAA MCMVVTRTEKDSYVVA F SDEM VPCPVTTDMTLQQVLM  
 AAF19049.1 YTTFKTV EPTGKRFLLA DVSASMMNQRALGSVNLNASTVAAA MCMVVTRTEKESSVVA FACDMVPPVTTDMTLQQVLT  
 AAC38001.1 YKSFKLV EPTGKRFLLA DVSASMMNQRVLGSILNASVAAA MCMVLVARTEKDSHMMVA F SDEM LPCPI TVNMLLHEVVE  
 CAA98241.1 YKAFVNAPPPTGKRYCLALDVSGSMTSRVSSSPLSCREAATGMSLINVHNEAEVRCVA FCDKLTELPFTKDWKIGQVND  
 P10155 AMSQIPAGGTDCSLPMIWAQKTNTPADVFIVFTDNETFAGGVHPAIALREYRKMDI-PAKLI VCGMTSNGFTIADPD  
 AAF19049.1 AMNKVPGNTDCSLPMIWAQKTTGAADVFI VFTDNETFAGQVHPAVALREYRKMDI-PAKLI VCGMTSNGFTIADPD  
 AAC38001.1 KMSDITMGSTDCALPMLWAQKTNAAIDFIVFTDCETNVEDVHPATALKQYREKMGII-PAKLI VCAMTSNGFSIADPS  
 CAA98241.1 YVNNLDFGRTDCGGLPMTWATENNLFKDFVFIYTNDNTWAGEIHPPFEAIKKYREA SGIHDAKVI VMAMQAYDYSIADPS  
 P10155 DRGMLDMCGFDTGALDVIRNFTLDMI  
 AAF19049.1 DRGMLDMCGFDTAALDVIRNFTLDMI  
 AAC38001.1 DRGMLD1CGFD SGALDVIRNFTLDMI  
 CAA98241.1 DAGMLD1TGFD SAVPQIVHEFVTGKI

470 480 490 500 510 520 530 540  
 550 560 570 580 590 600 610 620  
 630 640 650

40/40